

Sacrifice for the greater good?

Genome sequencers intend that their community and others should deposit data in community databases immediately, even if it risks the loss of publishing priority. But enforcement of this ideal could be a step too far.

Some US and UK genomics researchers are seeking to extend their principles of open access throughout the world of biology in unprecedented fashion. They claim that enforcing these principles would be in the best interests of science, and they may be right. But anybody believing that researchers in other disciplines and countries are ready to sign up would be wrong.

The principles concern databases — not just any databases, but those that are so widely used as repositories and sources that they are seen as ‘community resources’. A closed meeting in Fort Lauderdale last month explored the conflicts for the originators of genomics data between, on the one hand, the principles of immediate deposition and unconditional access, and on the other, the need to protect the originators’ rights to publish, and gain appropriate priority for, the outcomes of their labours.

This dilemma has been hanging over the community for some time, and has led to failed attempts to establish a set of ground rules to which the community could sign up. The problem has occasionally hit the headlines — for example, in a debate between the sequencers of the genomes of the malaria parasite *Plasmodium falciparum* and *Trypanosoma brucei*, which causes sleeping sickness (see *Nature* **405**, 601; 2000), and in conflicts over early use of the sequence data of the protozoan *Giardia lamblia* (see *Science* **295**, 1206; 2002). And the genome sequencers can cause problems themselves: sometimes they sit on data for an unreasonable period. The rat genome highlights another problem with immediate openness: the first pre-publication assembly has numerous errors, but is already deposited in public databases. Researchers who rush in and use the data may come to regret doing so.

A new attempt to establish ground rules, backed by the Wellcome Trust and the National Human Genome Research Institute (NHGRI, part of the US National Institutes of Health), is more radical than previous efforts, not only in the rules themselves but also in their enforcement and scope. So much is clear from proposals that form part of a report of the Fort Lauderdale meeting, due to be published this week on the Wellcome Trust’s website. They would remove all restrictions on the use of genome data.

Cause for concern

Deposition of data on an open-access website is, rightly, not considered to be equivalent to peer-reviewed publication. Under the new proposals, the originators of such data, if publicly funded, are in effect giving up any protection of their rights to claim priority in a publication. Anybody can download the data and publish whole-genome analyses, even if they scoop the originators. Attempts to impose licensing agreements to prevent such use (previously described by some as data misappropriation or, more bluntly, piracy) would henceforth be forbidden, the penalty being that centres attempting to impose such licences would be ineligible for funding.

The proposals are already causing concern in other countries (see page 877), where the principles of immediate and unrestricted access so successfully adopted by the international Human Genome Project are still not readily accepted. Even in Britain and the United States, not everyone is on board. After all, these principles are all fine and

dandy for the grandees of the field, who have little to lose in funding and reputation, but are tough on others who are less favoured.

The proposals are long on supporting data release, backed by funding-agency pressure, but short on any means of enforcing the other central requirement: that everyone should behave honourably and ensure that the originators of data get due credit. *Nature* is all too aware that under the pressure to publish, collegiality and seemliness frequently go out of the window.

To be fair, the proposals seek to compensate originators; they propose that genome sequencers publish statements of intent and a project description containing the scope of data and analysis that the originator expects to undertake. This would be a citable means of giving them credit in case they are scooped at the other end of the process. This is a good idea as far as it goes: one potential example has already appeared (M. V. Olson & A. Varki, *Nature Rev. Genet.* **4**, 20–28 (2003); doi:10.1038/nrg981). But will it compensate for the loss of incentive for researchers to stay in the business of data generation if they lose priority protection? Researchers now have an opportunity to feed in their views before the NHGRI’s advisory council considers the proposals for its endorsement in May.

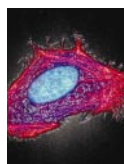
Broader community

Most mammalian sequencing projects have adopted principles of openness and unconditional access, whereas microbial and plant sequencers have been less forthcoming. The new proposals potentially project themselves beyond genomics: they suggest that any project where a ‘community resource’ is the objective should subject itself to the same principles of immediate release without conditions. The Fort Lauderdale meeting focused on genomics, so it remains to be seen whether other branches of biology data generation will accept the fundamental principle espoused: that science will benefit from immediate openness in advance of publication even if data originators lose out. But whatever the discipline, the idea that funding centres should insist on the adoption of such principles smacks of coercion, and may drive sequencers away from public funding.

What of the role of journals? Some referees have refused to review papers from groups that have not deposited data in databases at the time of submission. That is their right, but it is the editors’ job to steer around such indirect imposition of principles. *Nature* has received papers that might have scooped genome sequencers, but judged them to be inappropriate for scientific reasons; they were published elsewhere. We expect that to change — such situations will arise more and more often as bioinformatics improves.

Editors and peer reviewers must do what they can to ensure that sufficient credit is given to originators of data, and we will generally check papers with originators anyway to ensure that the data are reliably deployed. We will try to ensure that licensing agreements (for as long as they continue) are not contravened. And we will support originators’ rights to publish, and be generous in interpretation as to whether a previous publication has scooped them or not. But if a good piece of whole-genome analysis arrives on our desks, we’ll publish it whoever it comes from. That, as the Fort Lauderdale proposals also say, is in the best interests of science. ■

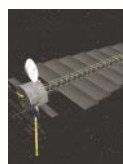




Cell division
Institute launched to
plunder life sciences
for space travel
p878



Cold draft
Polar organisms
line up for genome
sequencing
p880



The ion age?
NASA plans nuclear-
powered mission to
Jupiter's moons
p881



Deadly disease
Bird flu returns to
infect humans in
Hong Kong
p882

Draft guidelines ease restrictions on use of genome sequence data

Carina Dennis

Researchers working on large-scale publicly funded genome-sequencing projects must release their data as soon as possible, without imposing restrictions on the use of the information, say guidelines from two of the main funders of such work.

The US National Human Genome Research Institute (NHGRI) this week issued a draft revision to its policy on the use of data from large-scale genome projects. Existing rules prohibit users from publishing a whole-genome analysis before the sequencers' initial publication on the complete genome. "That restriction will be removed and will not be present in any other genome databases that we support," says the institute's director, Francis Collins.

The move follows discussions at a meeting of sequencers, bioinformatics researchers and journal editors last month in Fort Lauderdale, Florida, aimed at resolving disagreements over data access. The revision is expected to be approved by the institute's advisory council in May.

The Wellcome Trust, the London-based medical charity that convened the meeting, will also apply similar rules to the large-scale genome-sequencing projects for which it provides funding. Both organizations may also extend the rules to other large-scale



Francis Collins supports the publication of genome analyses before the sequence is completed.

ventures, such as protein-structure projects and analyses of gene expression.

The NHGRI's guidelines say users should acknowledge that "sequence producers have a legitimate interest in publishing peer-reviewed reports" about their sequence, and that "the entire scientific community can also help ensure that the system works fairly". They are an extension of the Bermuda Principle — the 1996 agreement between the

leaders of the Human Genome Project to make sequence data available without restrictions within 24 hours.

The move is likely to find favour with bioinformatics researchers, who use the sequence, but sequence producers may worry about being scooped. "Some groups may lack the integrity to play gentlemen's cricket," says Richard Gibbs, director of the Human Genome Sequencing Center at

Mixed results win HIV vaccine a guarded response

Hannah Hoag, Washington

Statisticians are urging caution over intriguing results from clinical trials of an HIV vaccine, which suggest that some ethnic groups may benefit from the drug.

Biotechnology company VaxGen announced the results of its AIDSVAX trials on 24 February. The overall results were disappointing, with only a 3.8% reduction in infection among those who were vaccinated compared with those who received a placebo. But the vaccine shows more positive results for certain ethnic subgroups, the company claims. Among black volunteers, for example, only 2% of those who received AIDSVAX

became infected with HIV, compared with 8% of those who received a placebo.

"The subset analysis showed surprising and provocative results in minorities," says Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases (NIAID).

But of the 5,009 participants in the study, only 314 were black, and some researchers warn against drawing premature conclusions. "Subgroup analysis has to be seen as more exploratory," says biostatistician Steven Self of the Fred Hutchinson Cancer Research Center in Seattle, Washington. "Over the next few months, the detailed analysis will show

whether there is an underlying biology that could explain the results."

VaxGen, which is based in Brisbane, California, plans to investigate factors such as age, sex and behaviour among the subgroups to see if something other than the vaccine can explain the results. It says that a more detailed analysis of the trial's results should be available by the end of March.

Several other possible vaccines may be less than five years away from large-scale clinical trials. The NIAID is developing two candidates for testing, and drug company Merck, based in Whitehouse Station, New Jersey, is also working on a vaccine. ■

► Baylor College of Medicine in Houston.

The team working on the malaria parasite *Plasmodium falciparum* was angered when other researchers tried to publish studies of sequence data before the project was complete (see *Nature* **405**, 601–602; 2000). But Collins believes this risk is a small price to pay to ensure the immediate release of data. “We will be scooped on a few occasions,” agrees Jonathan Eisen of The Institute for Genome Research in Rockville. “We’ll just have to deal with it.”

The guidelines could also hamper international collaborations. “While we as scientists are willing to release the data, the funding agencies in Japan may not support data release without restrictions,” says Yoshiyuki Sakaki of the RIKEN Genomic Sciences Center in Yokohama City, Japan. “Chinese agencies will face a dilemma,” adds Huanming Yang, head of the Beijing Genomics Institute. “Why should they invest the money if they can get data for free from the United States?”

But some question whether there will be a rush to publish ahead of the sequence centres. “I would much rather the annotation be done at the sequencing centre, rather than a third party who doesn’t know how the data were generated,” says Sean Eddy, a bioinformatician at Washington University School of Medicine in St Louis, Missouri. “We can’t be biting the hand that feeds us.” ■

Physicists fail to find saving grace for falsified research

Philip Ball

In the furore that followed last year’s revelation that Jan Hendrik Schön had fabricated data in a string of high-profile papers, one question remained unanswered: was there a nugget of truth in any of his reported findings? Two teams have now attempted to replicate some of Schön’s key results, and find that little of use can be salvaged from them.

The two groups recreated studies by Schön, who was based at Bell Laboratories in Murray Hill, New Jersey, on molecular field-effect transistors. The papers that they revisited featured examples of 7 of the 16 different types of data falsification or fabrication of which Schön was found guilty last October¹.

Field-effect transistors — electronic devices that can switch and amplify currents — are the workhorses of silicon circuits. Schön claimed to have observed transistor-like behaviour in a layer of an organic compound, just one molecule thick, sandwiched between two metal electrodes. This suggested that the size of transistors, and thus of integrated circuits, could be hugely reduced.

Teams of physicists at the Delft University of Technology in the Netherlands and IBM’s Thomas J. Watson Research Center in Yorktown Heights, New York, have now investi-

gated devices based on Schön’s design^{2,3}. Both groups followed the methods in his papers: the electrodes were made from thin layers of gold, between which rod-like organic molecules capped with gold-binding chemical groups provided a conducting pathway.

Neither group obtained the results recorded by Schön. Jeong-O. Lee and colleagues at Delft, for example, say that their devices were unreliable from the outset. Of more than 1,000 prepared, many were short-circuited — the current simply flowed between the two electrodes, heedless of the intervening layer of molecules. Only in 5–16% of devices did the current seem to flow through the molecular layer. Even then, the devices quickly degraded.

Coming from two of the world’s most experienced molecular-electronics labs, the results imply that the approach is fundamentally flawed. But the new studies have highlighted requirements for making similar molecular transistors, such as better methods for purifying the components. “There is no scientific reason to lose faith in the tremendous promise of the field,” insists Cherie Kagan, a member of the IBM group. ■

1. Brumfiel, G. *Nature* **419**, 419–421 (2002).
2. Lee, J.-O. et al. *Nano Lett.* **3**, 113–117 (2003).
3. Kagan, C. R. et al. *Nano Lett.* **3**, 119–124 (2003).

NASA seeks inspiration from microscopic views of life

Jonathan Knight, San Francisco

The space-shuttle programme may be on hold, but NASA researchers are still dreaming of the future. Ten days after the Columbia broke apart on re-entry, scientists gathered at the University of California, Los Angeles (UCLA), to discuss the development of futuristic spacecraft modelled on living cells.

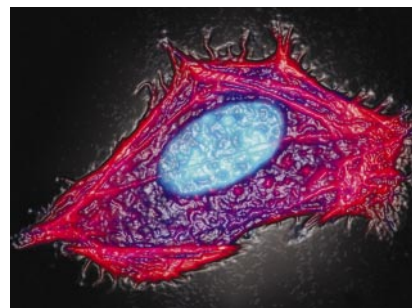
The day-long event, held on 10 February, inaugurated the Institute for Cell Mimetic Space Exploration, which is funded mainly by a ten-year, US\$30-million grant from NASA. The agency hopes that the institute’s 15 principal investigators, housed at UCLA and several other southwestern universities, will come up with biology-inspired devices that could facilitate space travel 30 years from now.

It’s an open-ended goal, admits Harry Partridge, an administrator at the NASA Ames Research Center in Moffett Field, California. “Instead of asking what do we need and how do we get there, this asks what is possible and what we can do with what we come up with,” he says.

Most of the research groups are exploring basic cellular processes that might later be scaled up. Biomedical engineer Carlo Montemagno’s group at UCLA, for example, is developing microscopic sacks of biological reagents — which the group calls ‘biobugs’ — that propel themselves, amoeba-like, across a substrate. The sacks will include growing filaments of actin, which normally make up the skeletons of living cells. According to Montemagno’s theory, the biobags will move at several micrometres per minute by extending an actin filament and pulling themselves along it.

Montemagno says that the goal is to endow the biobugs with the ability to sniff out and move towards specific substances, in much the same way that nerve tips grow towards chemical signals in the body. Biobugs could be dispatched en masse to search a spaceship for chemical or biological contamination, Montemagno suggests.

Other groups are copying different systems from the book of life, such as networks that gather sunlight, transmit



Cellular example: NASA researchers hope that cell biology will inform future space travel.

information through chemical signals and repair structural damage.

The inaugural event was planned before the Columbia disaster, and NASA administrators decided to hold it as scheduled. “In the spirit of exploration, things must go on,” says the director of the institute Chih-Ming Ho, an aerospace engineer at UCLA. ■

► www.cmise.ucla.edu

DAVID BECKER/SPL

Climate panel to seize political hot potatoes

Quirin Schiermeier, Munich

The politically sensitive issues of carbon sequestration and regional climate forecasts are to form part of the fourth assessment report by the Intergovernmental Panel on Climate Change (IPCC).

The assessment, due in 2007, was discussed by some 300 IPCC members in Paris on 19–21 February. Sequestration was chosen as the subject for a special report, separate from the main assessment. After the meeting, Rajendra Pachauri, director of the Tata Energy Research Institute in New Delhi and chair of the IPCC, confirmed that more detailed regional models of the impact of climate change would be considered by the assessment's authors.

Previous regional projections have generated controversy. A 2000 report from the US Global Change Research Program described the possible regional effects of climate change in the United States, but the results of this and other similar studies were deemed too unreliable by the Bush administration to be included in its strategy for climate-change research, released last November (see *Nature* **420**, 110; 2002).

"I am aware that there is an opportunity for much political debate when you start to predict the impact of climate change on specific regions," said Pachauri. "But if you want



Rajendra Pachauri: information is vital for action.

action you must provide this information."

Carbon-sequestration schemes, such as using empty oilfields to store the carbon dioxide from power plants, also provoke strong reactions. Environmental groups accuse advocates of the idea of damaging attempts to reduce emissions of greenhouse gases.

But the panel's fourth assessment will come too late to provide scientific input into negotiations on greenhouse-gas emission targets for 2013 onwards, the second phase of the Kyoto Protocol. Talks on these targets begin in 2005. "We will have to come up with some means to update negotiators with the latest scientific information," said Pachauri. One

possibility, he said, would be for some panel members to hold workshops to provide tentative directions for the negotiating parties.

The run-up to the Paris meeting was overshadowed by criticism of a previous IPCC special report on emission scenarios, which explored how changes in global economic and social conditions could affect emissions. In several papers over the past year, economists Ian Castles of the Australian National University in Canberra and David Henderson of Westminster Business School in London argued that the scenarios were "technically unsound" because of false assumptions on likely economic growth in developing countries.

Pachauri and others at the meeting agreed that the projections may need minor changes, but denied that they are flawed. "The emissions scenarios covered a very broad range," said Thomas Stocker, a climate modeller at the University of Bern in Switzerland, and a lead author on the third assessment report. "It is extremely unlikely that they do not include the actual development of emissions."

Details of the next assessment will be approved at an IPCC meeting in October. The IPCC will then appoint about 2,000 authors and reviewers from around the world, who could then begin work next year. Pachauri says he is keen to involve more young authors. ■

Experts cast doubt on Britain's green energy ambitions

Natasha McDowell, London

Plans for a new UK Energy Research Centre and a dedicated facility to test ocean-wave energy are the highlights of the research component of Britain's energy strategy, published on 24 February. But experts have questioned whether these and other proposals have the political and financial backing to meet the government's goal, announced in the strategy, of cutting carbon dioxide emissions by 60% by 2050.

The plans, unveiled in a government

white paper on Britain's energy needs, recognize energy-efficiency measures and renewable energy sources as the principal means of cutting emissions. The paper sets a goal of having 20% of Britain's energy derived from renewable sources by 2020 and includes no plans for new nuclear power stations.

"It made a lot of the right noises, but did not identify any new policies committing the government to these goals," says Paul Ekins, a professor of sustainable development at the University of Westminster, London.

Such scepticism is derived in part from a lack of concrete measures to help promote renewable sources. One exception is the publicly funded wave-energy test centre — the first in Europe — that will open this year off the coast of the Orkney Islands, Scotland. Costing around £5.5 million (US\$8.7 million), it will assess full-scale wave-energy machines. "The site is vital for the commercial development of wave energy," says Max Carcas of Ocean Power Delivery, an Edinburgh-based company that is developing a commercial offshore wave-power device.

Also highlighted in the white paper is a new UK Energy Research Centre, which will be established in the coming year. Details

are still being finalized, but the facility is intended to act as the hub of a new National Energy Research Network, bringing together the strands of energy research funded by the government. The funding for the centre is expected to be £8–12 million over five years.

These projects will be complemented by an extra £60 million in grants to expand the use of existing renewable technologies, such as wind power. But this will take government spending on such technologies to only £348 million over the next four years — a figure that some feel is too small to meet the new emissions goals. Jeremy Leggett, chief executive of Solar Century, a London-based company that sells solar panels, says that the funding is an order of magnitude less than in other European nations such as Germany.

Others say that by rejecting nuclear power and not investing enough in renewable sources, the government has dodged the question of how to move away from using oil and gas. "The white paper shows a lack of courage to make the hard decisions necessary to move this country away from dependence on fossil fuels," argues David Wallace, vice-chancellor of Loughborough University and vice-president of the Royal Society. ■



Wind of change: the UK government has pledged to increase Britain's use of renewable energy.

Genomes take pole position in the icy wastes

Hannah Hoag, Washington

Our knowledge of the Arctic and Antarctic has been transformed since the first successful expeditions to the poles nearly a century ago. Now, the US National Academy of Sciences has mapped out its strategy for advancing our understanding of these inhospitable regions further. Its report, which was released on 7 February, puts genomic sequencing firmly at the centre of this endeavour.

The report's authors say that the National Science Foundation (NSF) should establish a Polar Genome Science Initiative to understand, at the genomic level, polar organisms and how they interact with the environment. Possible benefits of such a project include fresh perspectives on the mechanisms of evolutionary adaptation, and the potential to uncover potentially useful genes.

The NSF is currently studying the report's recommendations. "We are excited by this," says Karl Erb, director of the foundation's Office of Polar Programs. "These really are new frontiers."

Polar research has so far lacked a genomics component, says William Mohn, a microbiologist at the University of British Columbia in Vancouver, Canada, who helped to prepare the report. "In the past, people have been doing physiological studies but they now need genomic information to do cutting-edge research," he says.

The report recommends that the NSF makes it a priority to assess which polar organisms are best suited to a wide range of scientific studies, and so should be selected for analysis. It also argues that non-polar researchers with expertise in genomics should be invited to participate in the project, and that international collaborations should be set up.

The results of a polar genomics initiative would be interesting, predicts William Detrich, a marine biologist at Northwestern University in Boston who chaired the committee that prepared the report. The polar regions offer unique opportunities to study evolution, he says, because of the way in which different organisms have adapted to the area's extreme conditions. Understanding this evolution could aid predictions of the impact that climate change will have on the region. Some fish, for example, have lost the ability to survive at temperatures above 4 °C.

The pay-offs might be even broader. The genes for molecules that allow polar organisms to survive low temperatures could, for example, be introduced into plants, increasing the amount of the world's land that is available for farming. Researchers also believe that some proteins produced by krill, shrimp-like creatures that live in polar seas, could find applications in food processing



In from the cold: krill are one of the polar organisms that may have their genome sequenced.

and storage. But before such applications can be developed, researchers must reveal the expression and function of the prospective genes and proteins.

To achieve this, better facilities will be needed. Limited equipment, poor infrastructure and disjointed Arctic and Antarctic programmes have impeded research, say the report's authors. They suggest establishing a

'freezer farm' somewhere in the United States so that researchers can gain access to polar organisms for pilot studies without having to journey to the poles. Detrich adds that polar researchers need better access to high-throughput sequencing facilities. The report suggests creating a virtual sequencing centre that would negotiate access to existing centres. ■

Canada boosts spending on science

David Spurgeon, Montreal

The farewell budget from Canadian Prime Minister Jean Chrétien, announced on 18 February, contains substantial spending hikes in many areas, including science.

As Chrétien enters his final year in government, Canada is the only country in the group of seven leading economic powers that is expected to record a surplus in 2002–03 — and most of that surplus will be spent in the new budget.

Research will benefit from a 10% increase — worth Can\$125 million (US\$84 million) — in the annual budget for the country's three federal granting councils, which allocate funding to university researchers. A further Can\$225 million will go to universities, colleges and research hospitals annually to help support this and other new research.

Part of this extra money will be channelled into the Polar Continental Shelf Project, which is responsible for providing infrastructure support for Arctic researchers. It will get a further Can\$6 million over the next two years. Polar researchers warned last month that this programme would disappear without a cash

injection (see *Nature* 421, 464; 2003). Other federal projects in the Arctic will get Can\$10 million over the next two years.

Chrétien has also boosted environmental technology. Canada ratified the Kyoto Protocol, the international agreement on limiting greenhouse-gas emissions, on 17 December, and the budget includes Can\$2 billion to help implement the government's plans on climate change.

But the wide variety of projects to be funded with this money, ranging from energy efficiency in buildings to fuel-cell projects, has left some observers fearful that infighting between government departments could dilute the impact of the new money. "We find the lack of detail very disconcerting," says John Bennett of the Sierra Club of Canada, an Ottawa-based environmental group.

A new programme will raise the number of government graduate scholarships by 70% to 10,000 by 2007. Administered by the granting councils, the scheme will support 2,000 masters and 2,000 doctoral students with annual awards of Can\$17,500 and Can\$35,000, respectively. ■

♦ www.fin.gc.ca/budtoce/2003/budliste.htm

NASA pins hopes on nuclear-powered orbiter

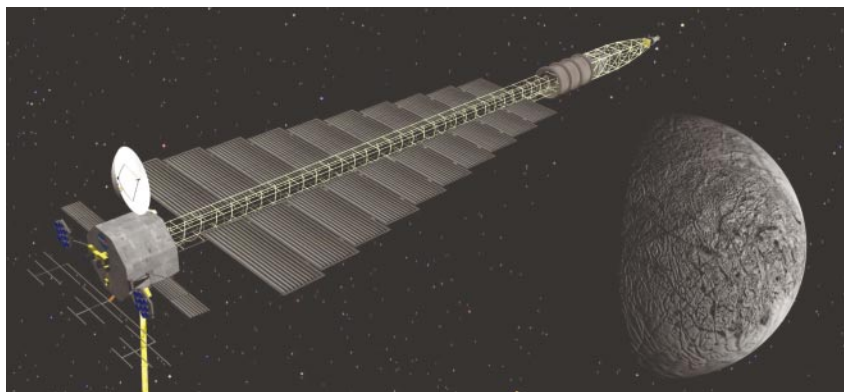
Tony Reichhardt, Washington

Daring, scientifically rewarding but fraught with technical and political risk. That's the way many space scientists view a proposed nuclear-powered Jupiter mission that would be the most capable — and probably the most expensive — planetary spacecraft ever launched by NASA.

The Jupiter Icy Moons Orbiter concept was unveiled earlier this month as part of NASA's budget request for 2004. Slated for launch no earlier than 2011, the orbiter is the centrepiece of Project Prometheus, a multibillion-dollar effort to develop nuclear power and propulsion for deep-space exploration. The planned budget for the orbiter alone would exceed \$90 million next year. Sources in Congress suggest that the total budget needed to launch the spacecraft could run to \$4 billion.

The orbiter's nuclear reactor would power an ion-drive engine: a device that expels ionized gas at high velocities to produce a small but continuous thrust. In weight terms, nuclear-powered thrusters are more fuel-efficient than those that use conventional chemical fuels. This would allow the orbiter to perform manoeuvres such as orbiting three of Jupiter's moons in succession. The orbiter is intended to circle three satellites — Europa, Ganymede and Callisto — for months at a time, seeking evidence of oceans beneath their icy surfaces.

Using a nuclear reactor also greatly



High flier: NASA's planned orbiter for Jupiter's moons could revolutionize planetary exploration.

increases the on-board power. The Cassini spacecraft, en route to Saturn, has about 875 watts of power on board. A nuclear reactor could produce tens of kilowatts. "This changes the entire regime of science experiments," says Colleen Hartman, head of the Solar System exploration division at NASA's headquarters in Washington. More powerful ice-penetrating radars could be flown and more data sent back to Earth.

NASA faces many hurdles if these benefits are to be realized. The agency has never launched a space-based nuclear reactor, for example. Keeping the reactor away from the scientific instruments may dictate a different shape for the orbiter, and NASA will have to learn how to stabilize and manoeuvre such a

spacecraft. Public protests have accompanied the previous launches of spacecraft carrying small amounts of nuclear material (see *Nature* **410**, 626; 2001), and the Columbia accident has enhanced fears over safety.

NASA will also have to convince Congress that the mission is worth \$4 billion. Advocates of planetary exploration say that nuclear propulsion is the best option, and that a similar system could be used on other missions if the orbiter is successful. But Louis Friedman, executive director of the Planetary Society in Pasadena, California, points out that if the project runs into problems, Europa exploration — a high priority for planetary scientists — could take longer than if conventional propulsion were used. ■

Europe draws up plans for funding agency

Marieke Degen, Munich

The future of science in the European Union (EU) began to take shape last week, as plans for a new basic-research funding agency gained momentum.

A high-level expert group met for the first time on 19 February in Paris, charged with outlining options for the creation of a European Research Council (ERC). The need for an independent, multidisciplinary agency has been intensively discussed during the past six months, after a meeting of scientists and research managers hosted by the Danish Research Councils (see *Nature* **419**, 108–109; 2002). Advocates argue that the council is needed to close the gap between the quality of basic science in Europe and in the United States. It would complement the EU's €17.5-billion (US\$18.8-billion) application-oriented Framework Programme.

Members of the ERC expert group, which consists of national research-council heads and other science administrators, left most

details unconfirmed, but agreed on basic principles. Delegates said that the council should not become political and should be committed to scientific excellence, rather than ensuring that each country receives as much back as it puts in.

The council's structure, size and funding mechanisms have yet to be defined. In particular, it is unclear to what extent any new agency will redistribute funds from existing EU and national research programmes, or how much fresh money it is likely to bring in. Discussions at the meeting about the council's budget covered figures from hundreds of millions to tens of billions of euros.

The idea has the strong backing of at least one section of the research community. Some 300 leading life scientists and representatives of national funding agencies held an independent meeting after the expert-group gathering, to discuss the planned council. Frank Gannon, director of the European Molecular Biology

Organization in Heidelberg, Germany, who helped to organize the second meeting, says that the council should focus on funding basic, investigator-driven research and supporting research infrastructure.

The expert group will refine its plans over the next few months and consult researchers outside the life sciences.

"It is important that scientists give us a feeling of what they need, and of which areas they think the council should cover," says Mogens Flensted-Jensen, vice-chairman of the Danish Research Councils and a member of the expert group. "But it is equally important to bring the ERC to the political agenda without further delay. Our task now is to work out a concrete proposal for how to launch it as soon as possible."

According to optimistic predictions, the council could come into being as early as 2004, when the Republic of Ireland takes over presidency of the EU, followed by the Netherlands. Both countries are expected to support the initiative. ■

Apocalyptic vision sees Newton book a date for doomsday

London Isaac Newton's laws of motion are among the most famous in science, but he also predicted something more sinister: that science, along with everything else, will stop when the world ends in 2060.

Newton's premonition has been unearthed in little-known manuscripts held in the Jewish National and University Library in Jerusalem. The pages reveal his attempts to decode the Bible, which he believed contained God's secret laws for the Universe.

Newton, who was also a theologian and alchemist, predicted that the second coming of Christ would follow plagues and war, and would precede a 1,000-year reign by saints on Earth — of whom he would be one. The most definitive date for doomsday, which he scribbled on a piece of paper, was 2060.

Newton's predictions will be explored in a documentary *Newton: The Dark Heretic* on BBC television on 1 March.

Columbia researchers sift through reclaimed results

Jerusalem Details of the results that can be salvaged from experiments on board the shuttle Columbia emerged last week. The craft was on a dedicated research mission when it disintegrated over Texas on 1 February.

Israeli researchers who had designed an experiment to study dust clouds blown from the Sahara Desert said that 51 minutes of a 77-minute video shot by the Columbia crew were transferred to the ground before the accident. Almost seven of the nine hours of data recorded on phenomena such as sprites — flashes of light that reach up above thunderstorms — were also beamed to the ground before the crash.

Other researchers, such as those working on studies of the radiation flux from the Sun and new spacecraft communication and navigation systems, say that they retrieved all

of their data before the accident. Most of the data from Columbia's life-science studies, however, were to be recovered on landing.

High-tech tomatoes get luxury residence

San Diego A complex of high-security greenhouses has become home to experimental vaccine-producing tomatoes at Arizona State University in Tempe.

The \$1.1-million greenhouse, which will prevent the release of pollen into the outside air, will house experiments to try to produce oral vaccines for Norwalk virus, cholera and the hepatitis B virus. The tomatoes grown in it have been genetically engineered to produce harmless proteins found in the viruses, which should prompt the body to develop antibodies against them. The tomatoes will be processed and freeze-dried, and the resulting material will be put into capsules for clinical trials.

The complex, which was made available for research last month, signals a growing interest in pharmaceutical crop research. Charles Arntzen, a plant scientist who directs the university's Arizona Biodesign Institute, says that collaborators are being sought for projects at the greenhouse.

Congolese health officials confirm Ebola outbreak...

Geneva Ebola virus has been confirmed as the cause of an outbreak of haemorrhagic fever in Kellé, a northwestern district of the Democratic Republic of the Congo. Eighty people have been taken ill and 64 have died in the country since late January.

The Congolese government has requested aid from the World Health Organization (WHO), and an assessment team has travelled to the site of the outbreak. "The people are very, very scared about what's happening, and they're looking for answers," says Iain Simpson, a communications officer for WHO in Geneva. The virus, for which there is no cure, causes external bleeding and kidney and liver damage.

Dead gorillas have also tested positive for Ebola in Mbomo, around 150 kilometres from Kellé. The virus affects non-human primates in much the same way as humans, but the natural reservoir from which it is contracted is still unknown.

...as bird flu sparks alarm in Hong Kong

Hong Kong Health officials are trying to isolate the virus responsible for two human cases of bird flu in Hong Kong. The flu is caused by a virus known as H5N1, which is usually found in chickens and aquatic birds, and there is no vaccine against it. Six people died in a 1997 outbreak in Hong Kong.



Hong Kong officials are trying to track a human outbreak of bird flu, usually found in chickens.

A 33-year-old man died of the virus on 17 February and his 9-year-old son is also infected. The boy's sister died at the weekend from flu-like symptoms, although it is not known whether she had been infected with H5N1. The family had recently visited Fujian Province in China.

One major concern is the possibility that the strain of virus responsible for the outbreak could jump between humans, although people who came into contact with the family have shown no symptoms of bird flu. Hong Kong health authorities and World Health Organization staff are studying the viruses to identify the source of the infection and so track the course of transmission.

Poorly pets to provide early warning in war on bioterror

Washington What's the best way of detecting the early signs of a bioterror attack? Keep an eye on people's pets, says Larry Glickman, an epidemiologist and veterinarian at Purdue University in West Lafayette, Indiana.

Next month, Glickman will begin trials of a surveillance system based on the electronic health records of the 60,000 cats and dogs treated every week at 300 branches of Banfield pet hospitals, which cover 43 US states. "Every night that information is processed," says Glickman. "The right programming and statistical analysis will allow us to detect a terrorist attack using agents such as anthrax or plague." The symptoms of both these agents can be mistaken for other common diseases such as flu.

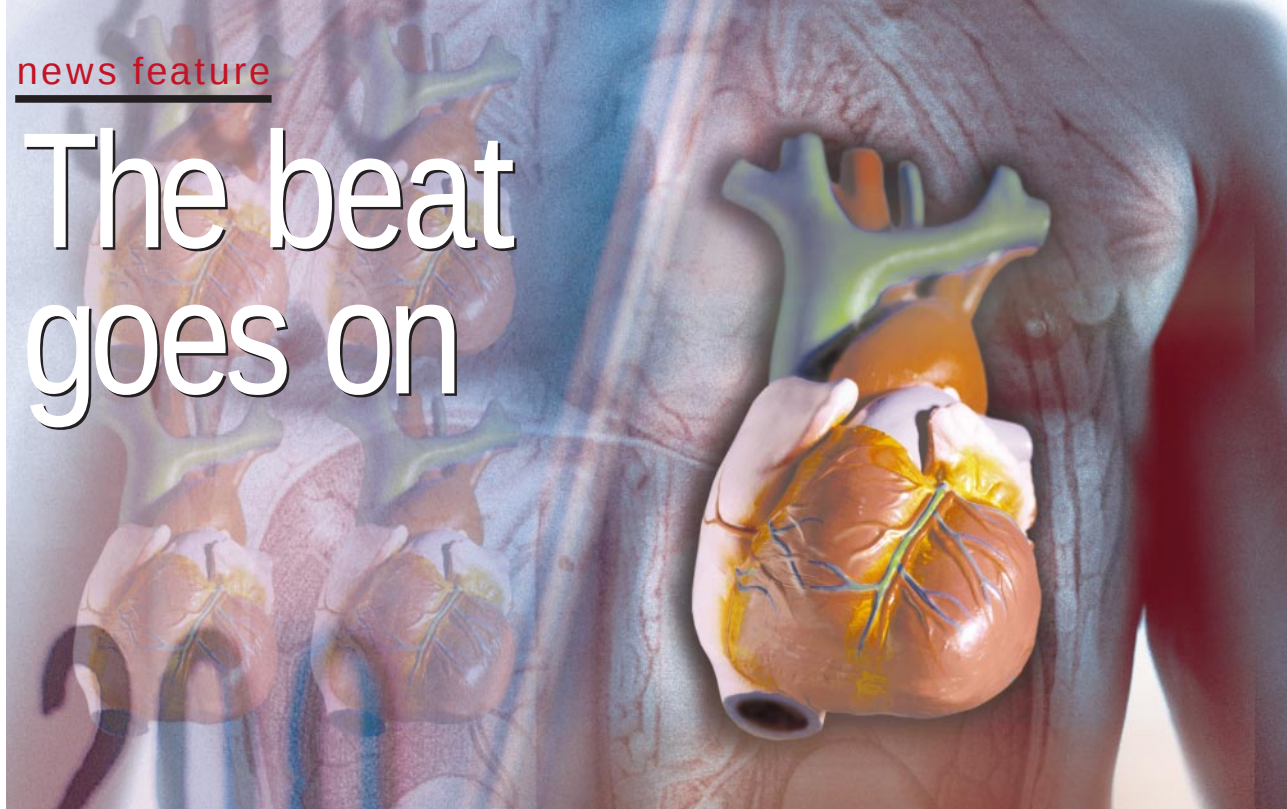
Similar analyses of human health records are being investigated as bioterrorism-detection measures (see *Nature* 421, 564; 2003), but Glickman says that human surveillance systems tend to be regionalized and use less standardized records.

NASA



Not in vain: researchers have salvaged some data from studies carried out by Columbia's crew.

The beat goes on



Four years ago, scientists claimed they would grow a functioning heart in the lab within a decade. That now looks like wishful thinking, but Catherine Zandonella finds that tissue engineers haven't given up on their grand vision.

Perhaps it was just millennial hubris, but in 1999 the world seemed full of enthusiasm for science and technology. The NASDAQ high-tech stock market was booming; the Human Genome Project was racing towards completion. And one group of scientists was talking up a plan to grow a fully functioning heart in the lab, starting with a culture of human cells in a petri dish. All they needed was ten years, plus about US\$5 billion, and our worries about the shortage of organ donors would be over.

The initiative, known as Living Implants from Engineering (LIFE), would rival the Human Genome Project in scale and in its potential to change medical practice. But four years later, LIFE has lost its lustre. The grand project remains largely unfunded, and functions mainly as a loose coalition of researchers who share their findings at meetings. "LIFE collapsed into the form of a club," admits its chief visionary, Michael Sefton of the University of Toronto in Canada.

Despite LIFE's stuttering start, tissue engineering retains enormous clinical promise. Tissue-engineered skin grafts have already won marketing approval from the US Food and Drug Administration, and implants of lab-grown cartilage are now in clinical trials. But these are relatively simple tissues. Growing organs as complex as the heart is an immensely greater challenge. "It would be wonderful if you could do it," says Robert Langer of the Massachusetts Institute

of Technology (MIT) in Cambridge, who helped to pioneer the field of tissue engineering in the mid-1980s.

'If' remains the operative word. "We are all dreaming of this artificial heart beating in a dish, but I think this is unlikely to happen over the next few years," says Thomas Eschenhagen, a tissue engineer at Eppendorf University Hospital in Hamburg, Germany. "The heart is simply too complex."

Growth area

Faced with this daunting complexity, several research groups are instead working on the interim goal of growing some of the heart's component parts — patches of cardiac muscle, valves, coronary blood vessels and so on. One of the largest projects is the BioEngineered Autologous Tissue (BEAT) initiative. Led by Buddy Ratner of the University of Washington in Seattle, BEAT received a \$10-million, five-year grant in 2000 from the US National Institutes of Health to create patches of cardiac muscle to repair the damage caused by coronary heart attacks.

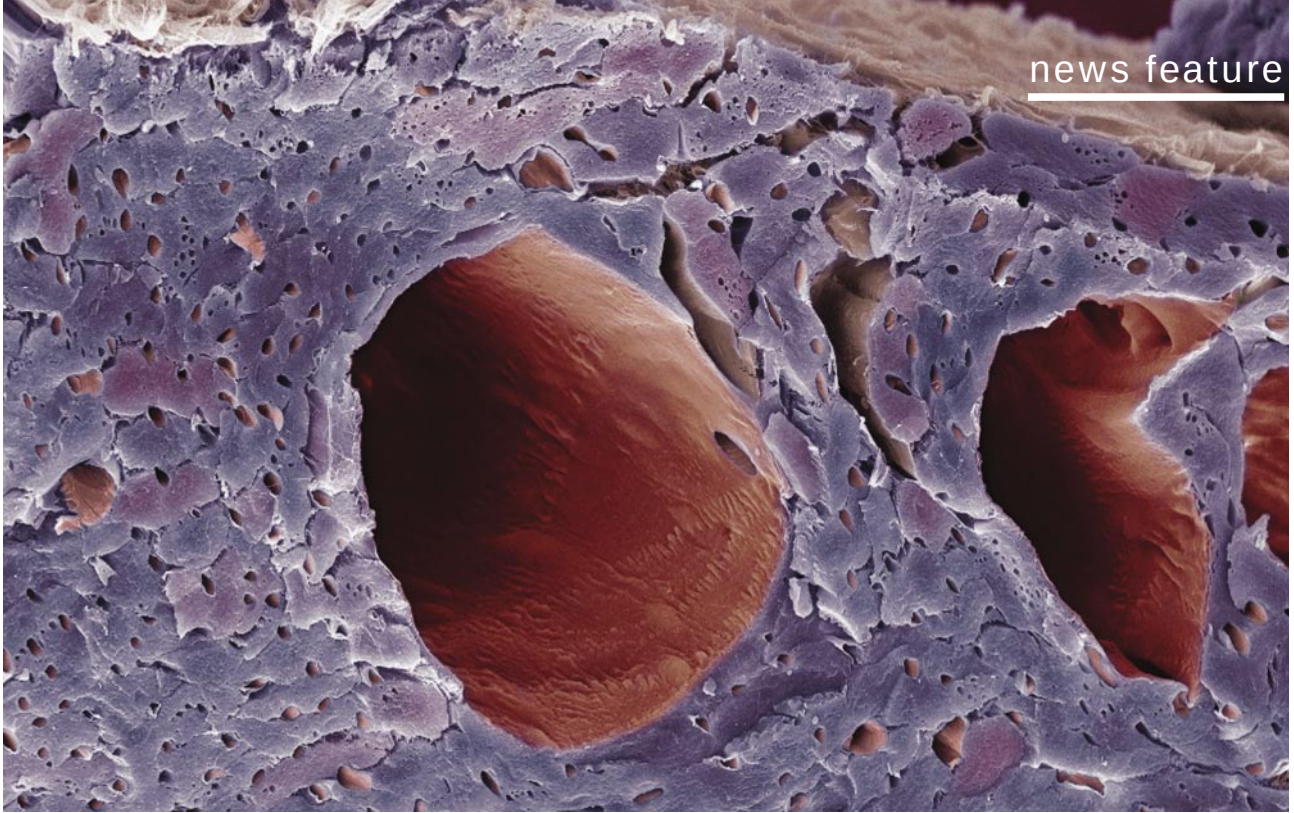
If that effort is successful, the BEAT team will try to create a complete ventricle, one of the muscular chambers that pump blood around the body. And in the long run, Ratner hasn't written off the idea of growing an entire heart from scratch. "The sheer audacity of the heart project keeps things exciting and stimulates us to think out of the box," he says.

There's also a pressing clinical need for

heart tissue. Presently, those who require a transplant can end up waiting several years for a donated organ, and their problems don't end there: because of immunological rejection and other complications, donated hearts often last for only a decade or so.

The basics of tissue engineering are relatively straightforward: first build a scaffold on which tissue can grow, then seed some cells into it, and tuck it away in a specially designed 'bioreactor' that provides nutrients and oxygen. When the cells have multiplied to fill the scaffold, remove the structure from the bioreactor and implant it into the body. In 1995, the principle was demonstrated to the world in those famous images of a 'human' ear made from scaffold-grown cartilage protruding from the back of a mouse.

Unfortunately, heart tissue is rather more difficult to grow. Adult cardiac muscle cells tend not to divide, so the scaffold must be seeded with cells from another source. One idea is to use skeletal-muscle cells, which proliferate readily. Over the past few years, researchers have also become excited by the potential of stem cells extracted from adult bone marrow, or from few-day-old human embryos, both of which can differentiate into cardiac muscle cells if given the right stimuli. Indeed, it may be possible to repair small areas of damaged cardiac muscle by injecting stem cells or skeletal-muscle cells directly into the tissue — ongoing clinical trials involving patients' own skeletal-muscle cells have



Heart of the matter: tissue engineers striving to grow replacements for natural cardiac muscle need to find ways to deliver nutrients to their tissues, much as capillaries (shown here in red) do in natural heart tissue.

yielded some improvement in muscle contraction. “We can use the heart itself as both a scaffold and bioreactor,” says Charles Murry of the University of Washington, a cardiac pathologist and a member of the BEAT team.

Although direct injection of cells could successfully mend small areas of damage, repairing the large areas of dead tissue that result when a major coronary artery becomes blocked will require the lab-grown patches of muscle that are BEAT’s main focus. To grow in three dimensions, cells need some sort of scaffold, which should be accepted by the body and ideally degrade over time into non-toxic components. The leading ideas involve using either biomolecules such as collagen, which provides mechanical support for natural tissues, or synthetic, biodegradable polymers such as poly(L)-lactic acid. To maximize the likelihood that seeded cells will ‘take’ to these scaffolds, some teams are coating their polymers with cell-adhesion proteins or designing them to release cell-signalling molecules¹. Eventually, researchers hope to create scaffolds that encode specific instructions for controlling tissue formation, by analogy to the signals that control embryological development.

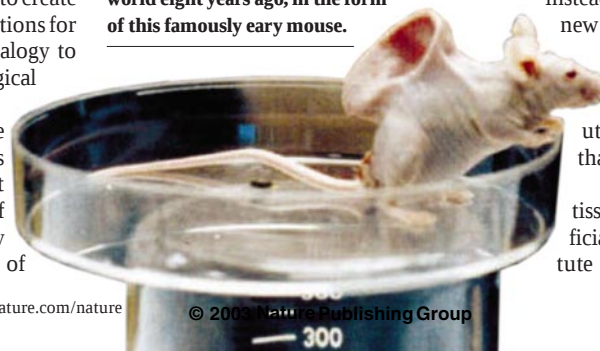
But there is still much to be done to perfect the mechanical properties of the scaffold. One of the most intriguing approaches is that of BEAT project member Kimberly Woodhouse of the University of

Toronto. Her group has designed a scaffold made of chains of elastin polypeptides, a biomolecule that provides elastic properties to some of our tissues. By experimenting with different molecules to crosslink the polypeptide chains, she can create structures that range from a delicate mesh to a material with the properties of an elastic band². For a patch of heart muscle, the ideal scaffold would be something that combines the properties of both — allowing cells to burrow into the matrix but providing as much strength and elasticity as a beating heart.

In thickness and in health

Perhaps the biggest challenge, however, is working out how to grow three-dimensional structures that contain more than a few layers of muscle cells. Most bioreactors simply cannot supply enough nutrients and oxygen to the growing tissue. Whereas human heart muscle is up to 2 centimetres thick, growth in a bioreactor typically stops once the tissue is about 100 micrometres, or 4–7 cell layers, thick. Beyond this thickness, the innermost cells are too far from the supply of fresh growth medium to thrive.

Listen up: tissue engineering was unveiled to the world eight years ago, in the form of this famously early mouse.



One approach is to sidestep the problem by abandoning scaffolds and growing single layers of cells that can then be sandwiched into slabs of tissue like a layered cake. Researchers led by Teruo Okano at the Tokyo Women’s Medical University are attempting this feat using cardiac-muscle cells from newborn rats — these cells are still able to divide. Okano’s team grows the cells on polymer surfaces that allow the intact cell layers to detach when the culture temperature is reduced³. In experiments reported last November at the Scientific Sessions of the American Heart Association (AHA) in Chicago, the researchers laid four of these sheets on top of each other until they fused. They then implanted them under the skin of immunodeficient rats. When the researchers opened the skin six months later, the engineered cardiac tissue was beating and blood vessels had permeated it.

Most researchers, however, are working on ways to simulate the presence of blood vessels in or around the scaffold to allow tissues to attain greater thicknesses. For instance, scientists led by Gordana Vunjak-Novakovic, a colleague of Langer at MIT, have designed a bioreactor that forces nutrient medium to perfuse directly through a cell-seeded scaffold instead of simply flowing around it. In new experiments using scaffolds containing cardiac-muscle cells from newborn rats, this bioreactor produced cardiac tissue with more evenly distributed cells and much higher cell densities than in any previous study⁴.

Eventually, it may be possible to grow tissues on scaffolds interweaved with artificial capillaries. At the Hope Heart Institute in Seattle, the BEAT project’s Robert

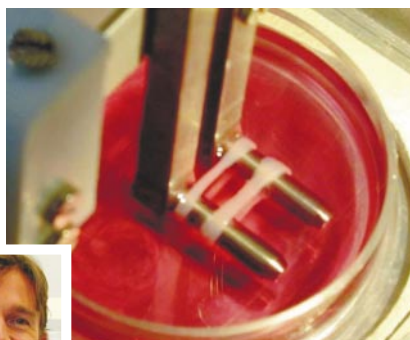
Vernon is designing one such artificial capillary bed. He spreads a dissolvable material out in thin, branch-like patterns, and then pours on a solution containing collagen and waits for it to solidify. Next, he uses a solvent to dissolve away the original material, leaving capillary-like tracks in the collagen. Vernon plans to sandwich sheets of this material between layers of cardiac muscle to create a three-dimensional block of tissue.

Vernon's vessels won't be able to carry blood because it would clot in the collagen tubes. But his artificial capillary beds could ferry cell-culture medium that could be crucial to keeping the tissue alive once it has grown. One of the challenges for the BEAT project will be in keeping the muscle patches oxygenated during implantation and for some weeks afterwards, until the patient's own vessels begin to colonize the implant. Vernon believes that his capillary bed could be hooked up to a miniature pump that would bathe the patch with nutrient fluid during this critical period.

Other researchers are experimenting with using the body itself as a bioreactor. In one such study, Jonathan Leor and his colleagues at Tel Aviv University's Sheba Medical Center in Israel seeded cells from embryonic and newborn rats onto three-dimensional scaffolds, cultured them in a regular bioreactor for four days to assess their viability, and then surgically implanted the tissue in rats' body cavities. While the tissues were in their rat hosts, blood vessels grew into them. After a week, the researchers removed the tissue and used it to patch holes they had made in the hearts of another set of rats. Last November, Leor told the AHA's Scientific Sessions that the tissue patches seemed to integrate successfully into the rats' heart tissue. "By growing tissues in the body cavity, it is possible to create an environment that duplicates the conditions for the development of a normal tissue," he says.

Growing a patch of heart muscle is only half of the battle: ensuring that it can actually pump blood is far from trivial. The heart is a powerful muscle, and engineered tissue needs to be similarly vigorous. So some researchers are sending their muscles to the gym for a workout. In Hamburg, Eschenhagen and his colleagues are making cylindrical patches of cardiac muscle using cardiac-muscle cells from newborn rats as a starting-point. But instead of seeding the cells on a scaffold, Eschenhagen mixes them into a collagen gel and casts them in a cylindrical mould. After a few days, he transfers the tissue patch to a stretching device that puts it through a workout that simulates the heart's contractions⁵. When Eschenhagen implanted the tissues into rats, that the cardiac cylinders contracted up to four times more vigorously than unstretched tissues⁶.

While Eschenhagen and others are per-



Take the strain: Thomas Eschenhagen stretches his engineered heart tissue to prepare it for the rigours of contraction.



fecting their techniques for growing cardiac muscle, other groups are trying to engineer other components of the organ. Finding a way to create replacements for blocked coronary arteries is important, because not all patients who need bypass surgery have sufficiently healthy vessels to provide their own grafts. A team involving Langer, for instance, has grown functional arteries by seeding a tubular scaffold with muscle cells from a cow's aorta. Once this had grown to fill the scaffold, the researchers seeded the tube's interior with endothelial cells, which form the smooth lining of blood vessels. Throughout the procedure, fluid was pulsed through the tubes to simulate the mechanical stresses that blood vessels must face. When implanted into miniature pigs, these tissue-engineered arteries seemed to be functioning well up to 24 days later⁷.

Hard graft

Cytograft Tissue Engineering, a company based in Novato, California, is taking an alternative approach. The company's scientists have grown sheets of tissue from human fibroblast cells taken from skin biopsies. They roll the sheets into tubes around stainless-steel cylinders and leave them to grow for several weeks, until the sheets fuse into their tubular shape. After removing the cylinders, they seed dog or rat endothelial cells on the inside of the tube to generate the vessel's inner lining. Finally, they expose the vessel to varying fluid flows and pressure to prepare them for implantation. In unpublished work, Cytograft scientists have grafted the resulting vessels into both dogs and rats, where they seemed to function well.



Michael Sefton: results in ten years?

Heart valves are especially tricky to grow because of their complex architecture of cells and extracellular matrix that continuously remodels itself to respond to fluid flow in the heart. Tissue-

engineered versions would be a boon for children who are born with defective valves, because mechanical implants must be successively replaced as the heart grows. A far better solution, in cases in which the problem is identified during pregnancy, would be to take a sample of fetal tissue and grow valves that would be ready for implantation shortly after birth, and which would subsequently grow along with the heart. To that end, John Mayer and his colleagues at the Children's Hospital Boston have seeded a biodegradable scaffold with endothelial and muscle cells taken from arteries in the necks of lambs. Again, the researchers pulsed fluid through the growing tissue to mimic the conditions experienced by a working valve. After implantation into the same lambs from which the cells were originally taken, the valves continued to grow, and survived for up to five months⁸.

Once valves, muscle, coronary blood vessels and other miscellaneous bits are constructed, the final step will be fitting the parts together into a functioning organ. Rather than growing each component separately and then tacking them together, the best approach would be to grow them all at the same time in a complex, multi-chamber bioreactor. But that's easier said than done. "The challenge will be to get them to grow in a coordinated fashion," says Frederick Schoen of the Brigham and Women's Hospital in Boston.

For the time being, it's a challenge that no one knows how to meet. But even if it proves impossible, tissue-engineered heart components might allow many patients who would otherwise need a replacement heart to carry on living without a transplant, thanks to careful running repairs. "Tissue engineering could allow damaged areas of the heart to be healed as the damage occurs," says Margaret Allen, a heart-transplant surgeon at the Hope Heart Institute, and a member of the BEAT team.

In retrospect, Sefton concedes that LIFE's ten-year timescale was unrealistic. "We were trying to capture the attention of the public," he admits. He now thinks that it may be 25 years before a patient receives a lab-grown heart, and refuses to speculate on the price tag. But he is still pursuing funding, and remains convinced that the project could produce tangible results within ten years. "We could probably have something that could work crudely," Sefton claims. "Not something ready to transplant into patients, but something you could hold in your hand." ■

Catherine Zandonella is a science writer in New York.

1. Hubbell, J. A. *Curr. Opin. Biotechnol.* **10**, 123–129 (1999).
2. Keeley, F. W., Bellingham, C. M. & Woodhouse, K. A. *Phil. Trans. R. Soc. Lond. B* **357**, 185–189 (2002).
3. Shimizu, T. et al. *Circ. Res.* **90**, e40–e48 (2002).
4. Radisic, M. et al. *Biotechnol. Bioeng.* **82**, 403–414 (2003).
5. Fink, C. et al. *FASEB J.* **14**, 669–679 (2000).
6. Eschenhagen, T., Didié, M., Heubach, J., Ravens, U. & Zimmermann, W.-H. *Transplant Immunol.* **9**, 315–321 (2002).
7. Niklason, L. E. et al. *Science* **284**, 489–493 (1999).
8. Hoerstrup, S. P. et al. *Circulation* **102** (suppl. III), 44–49 (2000).

Avalanche!

High in the Swiss Alps, a multidisciplinary group of researchers is developing a feeling for snow. Quirin Schiermeier meets the avalanche forecasters.

Many of those who lost their lives in Galtür on 23 February 1999 never knew what hit them. Until that afternoon, hundreds of thousands of tonnes of snow lay banked on the high ridges that flank the small ski resort in the Austrian Alps. Then the banks suddenly collapsed, unleashing an avalanche that swept away buildings and killed more than 30 people. Similar tragedies struck across the mountain valleys of Austria, Switzerland and France, where five metres of snow had fallen in the preceding weeks. In total, more than 100 people were killed, and the economic damage topped E1 billion (US\$1.1 billion).

The death toll for this disaster was exceptional, but avalanches kill every year. Only last month, 14 skiers and hikers died near Rogers Pass in British Columbia, Canada. A similar number have so far been killed in the Alps this winter. To counter the threat, studies of the layers of snow and the dynamics of avalanches have intensified in recent years. Researchers cannot eliminate the hazard but, by mapping the most dangerous areas and understanding the physics involved, they can issue timely warnings that could save lives.

Avalanches are a common phenomenon: each year about a million thunder down the slopes of snow-covered mountains around the world, from the Himalayas to the Andes. The vast majority do little harm. In fact, by

clearing patches of mountain forest they can provide space and light for new vegetation, and play an important role in maintaining the biodiversity of mountainous regions¹. But in the densely populated Alps, the events pose a threat to skiers, property and the tourism industry.

Lofty ambitions

The Swiss Federal Institute for Snow and Avalanche Research (SLF) in Davos is at the heart of the efforts to understand avalanches. Founded in 1936, the SLF is the oldest institute of its kind, and it is also the scientific backbone of Switzerland's avalanche-forecasting system. The SLF's staff love moun-



The Galtür avalanche killed more than 30 people.

tains. Many are semi-professional climbers with impressive expedition records. Summit photographs decorate most of the institute's offices and the talk is of recent climbs. The 125 scientists who try to make the Swiss mountains a safe place to live and relax also come from many different fields — geology, forestry, geophysics, soil science, meteorology and climatology. "There is no fixed way of becoming an avalanche scientist," says Paul Föhn, a geophysicist and the SLF's chief senior scientist. "They'll only learn what they need once they've joined us."

For researchers such as Föhn, understanding the behaviour of snow and avalanches is as great a challenge as the climbing expeditions they have undertaken. Most avalanches begin on slopes inclined at between 30° and 60° to the horizontal, when snowfalls produce 'snowpacks' that are about 0.5–1.5 metres high. But sometimes these packs reach 2–5 metres high, and when these collapse, they do so with devastating effects. The big question is how these snowpacks manage to remain stable for so long, allowing so much snow to build up.

The answer lies in the structure of the snow — crystals of ice that have air and water interspersed in the pores between them. The detail of this structure is far from straightforward, and it changes with time. The beautiful jagged crystals that make up

F. COFFRINI/AP

M. LEUKEL/REUTERS

fresh snow slowly become rounded, changing the way in which the snow is packed. Temperature also affects the structure. Banks of snow are warmer than the surrounding air, even warmer at their base, and all levels can be close to melting. "If you want to know how avalanches behave, you need to understand these complex physical properties," says Föhn.

Of all the factors involved, perhaps the most important for avalanche researchers are the interfaces between the various layers of snow that have fallen at different times. The top of a snowpack can form an icy crust known as hoar-frost, which forms a weak interface with any snow that falls on top of it. Under strain — the weight of the snow on top, a disturbance caused by skiers or snowboarders, or a sonic boom from a jet plane — the jagged crystals on the surface of the hoar, which mesh with the snow layer above, can suddenly collapse like a house of cards, sending the snow slipping away².

Catching the drift

But predicting when and where this will happen is extremely difficult. Perry Bartelt, head of the SLF's department of avalanche dynamics and simulation, has developed equations that describe the conservation of mass, energy and momentum in the snowpack³⁻⁵. These have been combined with meteorological data, such as measurements of relative humidity and wind speed, collected by the 50 or so observation stations in the Swiss Alps to produce a model known as SNOWPACK. This tracks the height of the different layers formed by successive snowfalls, as well as the nature of the interfaces between them.

On his computer screen, Bartelt shows how his model tracks the build-up of the snowpack in the Swiss Alps during the catastrophic winter of 1999 — it corresponds well with observations. But such simulations are plagued by the small-scale variations that are present in every bank of snow. An abrupt change in steepness in the middle of a slope may make the snowpack above it less stable, and SNOWPACK and other models cannot



Rolling, rolling, rolling: thousands of ping-pong balls simulate an avalanche on a ski jump.

take account of such infinitely variable details. Avalanche forecasters make use of the model's prediction for the height of the snowpack, but local knowledge is still essential. "Human observation and experience remain the most reliable components in any avalanche warning system," says Föhn.

Engineers working on avalanche defences and planning building projects would also benefit from a better understanding of snow. To calculate where an avalanche will come to a halt, and so prepare hazard maps, researchers use data on the frequency of avalanches in the area and on where the snow usually stops. This is combined with numerical models that describe how the mass of the avalanche changes as it falls — avalanches tend to pick up snow as they progress. To test the result of these calculations, researchers need to study avalanches under controlled conditions.

This isn't as contradictory as it seems. On 31 January, after a three-year break caused by lack of snow, SLF researchers used a small dynamite charge to release a powerful avalanche at a test site in the Vallée de la Sionne in Switzerland. Such experiments can prove hairy for the scientists involved. "Even in our small observation bunker at the foot of the slope we felt the enormous shock wave," says Felix Tiefenbacher, a physicist who studies avalanche dynamics. During a similar experiment four years ago, the 20-metre-high measurement pylon erected in the avalanche lane was torn down — although this time it withstood the pressure.

Data from sensors on the pylon and elsewhere yielded measurements of velocity, friction between the snow and the ground, and the pressure exerted by the avalanche. The researchers also measured the mass of snow that arrived at the base of the slope — for comparison with the mass used to trigger the avalanche — and how far the snow spread. In the long term, says Tiefenbacher,

they hope to use such data to generate a law to describe the flow of snow, similar to the fluid-dynamics equations that are used to describe the movement of water. Such a law could then be used to decide which areas are suitable for building, and how to protect existing property in danger areas. It may also aid research into similar processes, such as rock falls, landslides and mudflows.

Sharp chutes

Adverse weather conditions, and the risk of uncontrolled avalanches, mean that the large-scale experiments needed for this effort are not performed very often. In the meantime, researchers have to make do with substitutes. Tiefenbacher, for example, uses a 34-metre-long snow chute on the Weissfluhjoch, a mountain near Davos, to simulate avalanches and test the architecture of catching or deflecting dams. It is a painstaking business. "It takes hours to shovel into the chute the ten tonnes of snow you need for a two-second experiment," he says.

To estimate the balance of forces in an avalanche, researchers have even tried using ping-pong balls. In several experiments since 1995, researchers at Hokkaido University in Japan have released up to 550,000 balls from close to the top of the landing slope of the Miyanomori ski jump in Sapporo⁶. Video cameras measured individual ball velocities, and air-pressure sensors were mounted at different heights. Avalanches are much more complex events than such simulations, but the results helped to confirm assumptions, based on laboratory experiments, about the frictional forces and turbulence involved in such granular flows.

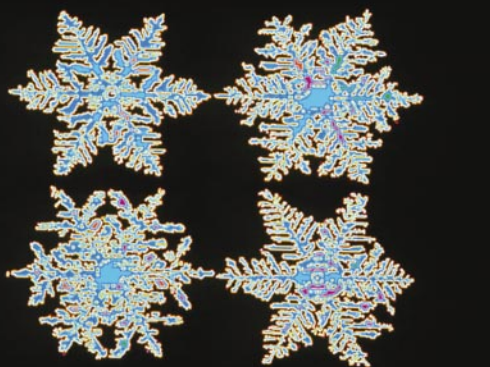
Despite the difficulties involved in their field, researchers at the SLF and elsewhere can point to a steady reduction in avalanche casualties. More than 100 people suffocated in their homes in the winter of 1950–51, trapped by avalanches. But apart from exceptional incidents such as Galtür, the main threat is to tourists who ignore warnings. The Davos scientists nevertheless remain alert to the constant winter threat. But they also get the chance to indulge in more glamorous work — the runs for this month's Alpine World Ski Championships, held in St Moritz in Switzerland, were prepared with the help of SLF expertise. ■

Quirin Schiermeier is Nature's German correspondent.

1. Johnson, E. A. *Ecology* **68**, 43–53 (1987).
2. Bader, H.-P. & Salm, B. *Cold Regions Sci. Tech.* **17**, 287–300 (1990).
3. Bartelt, P. & Lehning, M. *Cold Regions Sci. Tech.* **35**, 123–145 (2002).
4. Lehning, M., Bartelt, P., Brown, B., Fierz, C. & Satyawali, P. *Cold Regions Sci. Tech.* **35**, 147–167 (2002).
5. Lehning, M., Bartelt, P., Brown, B. & Fierz, C. *Cold Regions Sci. Tech.* **35**, 169–184 (2002).
6. McElwaine, J. & Nishimura, K. *Ann. Glaciol.* **32**, 241–250 (2001).

Swiss Federal Institute for Snow and Avalanche Research
www.slf.ch

S. CAMAZINE/SLF



Crystal balls: the physical structure of snow is key to predicting the behaviour of avalanches.

Switzerland's role as a hotspot of type specimens

Up to a quarter of all the world's species can be found in one small, landlocked country.

Sir — Switzerland covers 0.03% of the world's land surface, and lacks marine or tropical ecosystems. Yet we have found an estimated 345,000 type specimens (used for taxonomic naming of species) in Swiss museums and other collections. According to some estimates, there are about 1.5 million known species in the world. If correct, this means that Switzerland contains type specimens of up to a quarter of all the world's known species.

This exceptional number is surprising, as Switzerland has never been a colonial power, nor does it contain many large natural-history collections like those kept in prestigious institutions elsewhere. The work was done by taxonomists who built up enormous global collections — for example, Augustin-Pyramus de Candolle and his son Alphonse in the nineteenth century, and Auguste Forel, who died in 1931 — and by others who focused on the biological survey of particular

areas, such as New Caledonia or Paraguay.

Establishing the exact number of types in Swiss collections is part of an initiative by the Swiss Biodiversity Collections Online Consortium (www.biodiversity.ch/sbc-online.ch) — a consortium of Swiss systematists, curators of biological collections, the Systematics Task Force and the Swiss Biodiversity Forum. The consortium is lobbying the Swiss federal government to invest in this unique resource so it can be made available to the scientific community, with access to visual sources and databases, through the Global Biodiversity Information Facility (GBIF) and other Internet applications.

The type collections are invaluable tools for use in international efforts to document our planet's biodiversity. Not only should they stimulate taxonomists to work on them, but they should surely be important enough for Switzerland to become a full member of GBIF so it can share its wealth

with other scientists and conservationists.

This activity may revitalize Swiss national systematics research, which has suffered several recent setbacks. Currently, there are no professors of systematic zoology at any Swiss university, and the country has only two institutes of systematic botany (both with worldwide herbaria), in Geneva and Zurich. It will also provide an excellent incentive to tempt leading foreign scientists to work in Switzerland.

Donat Agosti

American Museum of Natural History, New York, New York 10024-5192, USA

Peter Linder Institute for Systematic Botany, University of Zürich, 8008 Zürich, Switzerland

Daniel Burckhardt Naturhistorisches Museum, 4001 Basel, Switzerland

Sylvia Martinez Swiss Biodiversity Forum, Swiss Academy of Sciences, 3003 Bern, Switzerland

Ivan Löbl Muséum d'Histoire Naturelle, 1211 Geneva, Switzerland

Pierre André Loizeau Conservatoire et Jardin Botaniques, 1292 Chambésy/Geneva, Switzerland

Novel and conventional paths to a better banana

Sir — Your News Feature “A dying breed” (*Nature* **421**, 568–570; 2003) highlights the threat to commercial banana plantations from a strain of Panama disease. Scare stories predicting the imminent end of the banana if this fungal pathogen spreads to Latin America have been blown up out of all proportion in the media (see, for example, *New Scientist* 26–29, 18 January 2003). Contrary to what has been written, the Panama pathogen with the ability to kill Cavendish (the main commercial subgroup of banana) in the tropics is found only in parts of a few countries, notably Malaysia and Indonesia.

Controls on the international movement of banana planting material should prevent this particular type of Panama from rapidly expanding its distribution. Even if this pathogen strain appears in plantations in Latin America, strict quarantine measures should limit its effects. Neither the export banana, nor any other banana, is in danger of disappearing just yet.

Because of the sterility of Cavendish cultivars, genetic manipulation is seen as the answer to disease problems that afflict this important crop. Although Cavendish is sterile, to label all bananas “sterile mutants” is an unkind exaggeration. Many clones are fertile and have been used in conventional breeding programmes.

New bananas have been successfully created by crossing synthetic diploids (derived from wild species with resistance to disease and pests) with cultivars in need of improvement. Much work has been undertaken at the Honduran Agricultural Research Foundation (FHIA) in the past, and its hybrids have found niches in various countries around the world. Many thousands are grown in Cuba, where chemical disease control is an unaffordable extravagance.

Certainly, the goal of breeding a disease-resistant replacement for Cavendish, the banana commonly seen in supermarkets, by conventional means has been elusive. However, the development of a new export banana is by no means impossible. The late Phil Rowe hoped to achieve this at FHIA by crossing fertile dwarf variants of ‘Gros Michel’, the old cultivar of the trades, with his improved diploids. More funding, which is sadly lacking, could still help his dream to be realized.

As reported with other crops, limited funds for banana improvement seem to have been channelled into genomic and genetic transformation work. If consumer objections to genetically modified food can be overcome, this research may have long-term benefits. However, we need to keep all our options open and also support conventional breeding methods.

David Jones

Editor, Diseases of Banana, Abacá and Enset, 12 Charlotte Brontë Drive, Droitwich Spa, Worcestershire WR9 7HU, UK

Objective assessment of transgenic salmon

Sir — In your News item, “Transgenic salmon still out in the cold in the United States” (*Nature* **421**, 304; 2003), you state that Aqua Bounty Farms is working with us “to perform an environmental-impact assessment, which it will use to support its application” to the US Food and Drug Administration for approval of its technology. Although we are discussing such work with Aqua Bounty, we have not yet signed a contract to do this research. More important, we understand that Aqua Bounty intends to submit our analysis as part of its application, but we do not know whether our results will support that application.

Our analysis will be a rigorous, objective, scientifically sound investigation, subject to standard legal separation between academic researchers and their funders. The work will be conducted in collaboration with a panel of scientific experts. Our results will not be shared with Aqua Bounty or regulatory agencies until submitted for publication in a scientific journal. Further, aware of the controversial nature of this issue and to inform the questions we ask and our analysis, our research will receive input from multiple funders with a range of perspectives on the topic of transgenic salmon.

George Gray

Harvard Center for Risk Analysis, 718 Huntington Avenue, Boston, Massachusetts 02115, USA

Liability for climate change

Will it ever be possible to sue anyone for damaging the climate?

Myles Allen

As I write this article in January 2003, the flood waters of the River Thames are about 30 centimetres from my kitchen door and slowly rising. On the radio, a representative of the UK Met Office has just explained that although this is the kind of phenomenon that global warming might make more frequent, it is impossible to attribute this particular event (floods in southern England) to past emissions of greenhouse gases. What is less clear is whether the attribution of specific weather events to external drivers of climate change will always be impossible in principle, or whether it is simply impossible at present, given our current state of understanding of the climate system. The issue is important as it touches on a question that is far closer to many of our hearts than global sustainability or planetary survival — who to sue when the house price falls?

At the heart of the problem is the distinction between weather and climate. As Edward Lorenz put it, “climate is what you expect, weather is what you get”. In the twenty-first century, climate is what you affect, weather is what gets you. Climate means ‘possible weather’, or what a statistician would call the ‘expected weather’ and its variability for a particular time of year, given all the properties of the ocean–atmosphere system, current levels of greenhouse gases, solar activity, and so on. The ‘attribution problem’ for externally driven changes in climate (as opposed to specific weather events) boils down to questions such as: “what would the climate have been like had we not increased greenhouse-gas levels?” This is a well posed question to which, if we define climate rigorously to encompass all the properties of the ‘attractor’ of atmospheric and oceanic weather, there is only a single answer.

In practice, all we can ever observe directly is weather, meaning the actual trajectory of the system over the climate attractor during a limited period of time. Hence we can never be sure, with finite observations and imperfect models, of what the climate is or how it is changing. This uncertainty can nevertheless be rigorously quantified, allowing formal probabilistic attribution statements. For example, the recent Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC TAR) concluded: “most of the observed warming over the past 50 years is likely [meaning, specifically, a better than two-in-three chance] to have

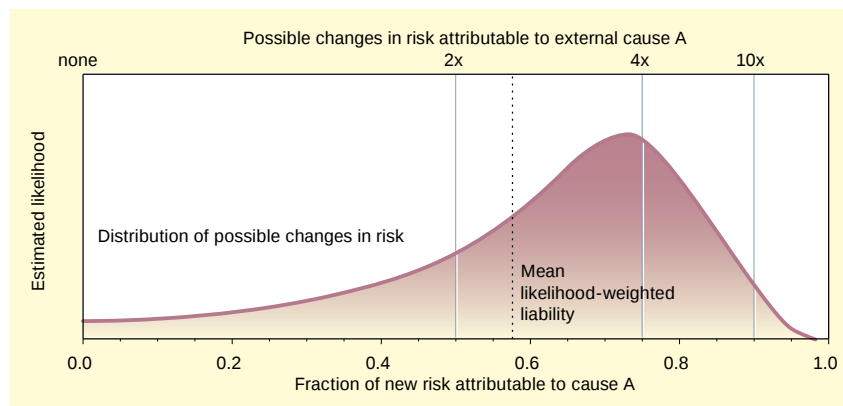


Figure 1 How we might be able to calculate liability for climate change. We will never know exactly how external drivers of this change, such as greenhouse-gas emissions, alter the risk of undesirable events, such as floods, but this does not prevent us working out a ‘mean likelihood-weighted liability’ by averaging over all possibilities consistent with currently available information.

been due to the increase in greenhouse gas concentrations.”

Not only is climate difficult to observe, but in most cases it is weather events that actually do the damage—for my neighbours in Vicarage Road, the information that we have a rigorous attribution procedure for changes in an unobservable attractor may seem of limited interest. Their feelings are running high, so let me explain at once that the science of detection and attribution may nevertheless have something to offer them. In a perfectly efficient and well-informed insurance market, premiums for flood-risk cover should be determined by the risk of flooding, which is a property of the climate, not the actual weather in any particular year (this assumes that seasonal forecasting never reaches a level of accuracy where insurers can adjust their premiums in the light of the forecast for the coming winter). So if insurance premiums rise as insurers factor in the increased risk of flooding due to climate change, and house prices consequently fall, some of this loss can straightforwardly be blamed on past greenhouse-gas emissions.

But how much? Any compensation settlement would have to define what fraction of a given loss was due to human influence

Even the most impassioned eco-warrior has nothing on a homeowner faced with negative equity.

on climate, and what fraction might have happened anyway, or happened for other reasons (for example, farmers upstream reducing the water-carrying capacity of their land). Of course, there are an infinite number of answers to this question, depending on the level of confidence required. Attribution statements, however rigorous, always need to be qualified with some level of probability, such as the two-in-three chance quoted above. As there can be no such thing as fuzzy compensation, how could such evidence be used in a settlement?

One approach to this problem is illustrated (Fig. 1). The curve shows how some external driver of climate change, such as past greenhouse-gas emissions, may have increased the risk of an undesirable event, such as the floods in Vicarage Road. There will always be some uncertainty in attributing changes in risk to external causes — in this schematic example, our ‘best guess’ is that cause A has increased the risk of this event by a factor of three, but there is still a 10% chance that it has not increased the risk at all. The lower axis shows the fraction of the new risk of this event that can be attributed to A — specifically, the amount by which current risk levels would be reduced if A were absent. If A has trebled the risk over its ‘pre-industrial’ level, then there is a sense in which A is ‘to blame’ for two-thirds of the current risk. To compute a single figure as a basis for compensation, one could simply average over all possibilities to give a ‘mean likelihood-weighted liability’, which in this case is somewhat less than two-thirds because of the uncertainty in how much the risk has increased (the spread of the distribution).

Would the concept of averaging over

possibilities, which is familiar enough in quantum mechanics, ever be acceptable to 12 honest citizens on a jury as a basis for a compensation settlement? This is not a scientific issue, but a legal one — the figure simply shows what science could, in principle, deliver. We are not yet in the position to produce such figures for the contribution of greenhouse-gas emissions to the increased risk of flooding in south Oxford (Fig. 2). But the point is, if we get the science right, we could be.

Who pays?

Figure 1 applies directly to damages related to changing risks rather than to actual events — essentially, the increased cost of insurance. But what about uninsured losses, or the losses incurred by the ultimate re-insurer of flood damage? We will never be able to say, at any confidence level, that human influence has contributed $x\%$ to an actual weather event. What we can say is that past greenhouse-gas emissions are likely (at some pre-specified confidence level) to have increased the risk of that event over its pre-industrial value. This does not preclude compensation settlements — juries have not been perturbed by the possibility that an individual smoker might in any event have contracted cancer. From a naive scientific perspective, however, what is the equitable solution?

Equal treatment of insured and uninsured losses suggests we should simply apportion liability according to the change in risk. If, at a given confidence level, past greenhouse-gas emissions have increased the risk of a flood tenfold, and that flood occurs, then we can attribute, at that confidence level, 90% of any damage to those past emissions. Again, we simply have to average over all possibilities consistent with current knowledge to arrive at a net likelihood-weighted liability. So, if courts can accept the concept of averaging over possibilities to produce an equitable distribution of liability, in theory, one day

Could current greenhouse gas emitters be held liable for the actual impacts of their emissions?

people driving up the local hill in their SUVs might be contributing to the cost of replacing the floors in Vicarage Road.

Some climate-change-related lawsuits have already been filed, but so far these have focused on technical legal issues, such as whether an adequate environmental assessment must cover climate change. The big question is whether current greenhouse-gas emitters could ever be held liable for the actual impacts of their emissions. The prospect of a class-action suit with up to six billion plaintiffs and an equal number of defendants may seem rather daunting, but if we can overcome these problems in end-to-end attribution, everything else is (at least conceptually) straightforward. Carbon dioxide is a well-mixed greenhouse gas, so an equitable settlement would apportion liability according to emissions, with some discounting over time to allow for the lifetimes of carbon dioxide anomalies in the atmosphere.

There are, of course, very substantial practical challenges in tracking down who has emitted what, but by the time we finish paying off our home loans in the early 2020s, almost two-thirds of greenhouse gases in the atmosphere will have been emitted post-1990 (the usual benchmark date at which climate change began to be considered a serious issue). So we could agree an amnesty for pre-1990 emissions without significantly affecting the final outcome, thereby avoiding the ethical dilemma of holding people liable

for emissions made before climate change was on the agenda. There are also challenging issues about where liability lies — with a company for selling fuel, or with an individual for driving? Such problems are not unique to climate change. If I were to sell you deodorant contaminated with dioxins, you would be unimpressed by the defence that the chemicals were doing no harm to anyone while in the can. Does the same logic apply to fossil fuels?

Help from the hidden hand

Opponents of such a class-action suit would doubtless argue that the consumer always pays in the end, and any initiative in this direction would inevitably increase the cost of fossil-fuel-based products, damaging the economy, jobs and so forth. Crucially, however, the size of the 'climate-change risk premium' would be determined by the hidden hand of the market, not by politicians in tortuous intergovernmental negotiations. There would no longer be any need to forge a near-global consensus on the risks of climate change before we agree on what to do about it. A market would emerge in cover against climate-related law suits. Companies that genuinely subscribe to the optimistic view that any climate change will be small and benign could decline such cover, but in an efficient market, they would then pay a corresponding premium on their cost of long-term capital. Paying extra for fuel to cover the cost of the oil company's climate risk insurance might feel rather like a carbon tax, but there is no point in blockading roads over it.

Many of the possible losers from climate change, such as polar bears or the inhabitants of the Earth in 2200, would be unable to benefit from any class-action suit that had any bearing on decisions made now. Their only protection is our collective environmental conscience, presumably expressed through government intervention. We all care about polar bears, of course, but to judge from the current rate of progress in post-Kyoto negotiations, it appears that we are not thought to care that much. And even the most impassioned eco-warrior has nothing on a homeowner faced with negative equity. Politicians have many things to worry about, so perhaps it is time to consider some apolitical mechanisms for redistributing the costs of climate change. First, however, there is science to be done — the impression held by much of the scientific community that the attribution problem for climate change is largely done, apart from mopping-up operations. That is certainly not the view in south Oxford. ■

Myles Allen is in the Department of Physics, University of Oxford, Oxford OX1 3PU, UK.
e-mail: myles.allen@physics.ox.ac.uk. He is writing here purely in his capacity as the chap at number 73 who was after some sandbags on Saturday 4 January.



Figure 2 A duck's delight: widespread flooding in January caused misery in Oxford.

Feeling emotional

What can a seventeenth-century philosopher possibly tell us about emotion?

Looking for Spinoza: Joy, Sorrow and the Feeling Brain

by Antonio Damasio

Harcourt/Heinemann: 2003. 368 pp. \$28/£20

Ray Dolan

In the past decade, Antonio Damasio has published two hugely influential books on human emotion, *Descartes' Error* (Picador, 1995) and *The Feeling of What Happens* (Harcourt, 1999). It is no coincidence that their publication coincided with a belated recognition, within psychology and the neurosciences, that emotion constitutes a fundamental component of mind. Both works had the enviable and elusive quality of appealing in equal measure to scientific and general readerships. *Looking for Spinoza* effortlessly embraces these qualities and, with its beguiling narrative (suggested in the title), provides the reader with an intimate account of intellectual insight and discovery.

Damasio's work on emotion has championed an important distinction between emotion and feeling. Emotion refers to the external consequences of processing an emotion-eliciting object; feeling refers to a private internal experience engendered by an emotional occurrence. Feeling states, which are given an extended treatment in this book, are closely bound up with neuronal mappings that index perturbations in the homeostatic state of an organism.

Damasio has suggested that these mappings provide the underpinning for an emergent self, an idea that he develops in finely measured tones in the opening passage of *Looking for Spinoza*. He suggests that human experience is founded in "feelings of myriad emotional and related states, the continuous musical line of our minds, the unstoppable humming of the most universal of melodies". One is reminded here of W. H. Auden's remarks on the passing of his fellow poet W. B. Yeats: "The current of his feeling failed; he became his admirers."

Damasio rediscovered the seventeenth-century philosopher Baruch Spinoza when verifying the accuracy of a long-cherished quote he had retained from Spinoza's *The Ethics*. What was intended as a brief diversion took on the quality of a journey. On re-reading Spinoza, Damasio experienced a deep resonance with his own ideas. Spinoza wrote extensively, and insightfully, about emotion. For Spinoza, any emotion was a variant of one of three primary affects, involving joy, sadness or desire. These emotions constituted universal properties of mind or common notions. Indeed, his assertion that "each one from his own affect,



Of one mind: Baruch Spinoza's view that the mind is ruled by the body struck a chord with Antonio Damasio (right).

judges, or evaluates, what is good and what is bad" presages recognition that emotions have a role in sensing value. Even more radical was his proposal that "the idea constituting the human Mind is the Body", which echoes Damasio's thesis that the very foundation of consciousness is bound up with representing bodily states.

Homeostasis and its regulation is a recurrent motif in *Looking for Spinoza*. In higher organisms, the biological imperative for self-preservation involves a complex of hierarchically arranged regulatory systems.

Damasio argues that a nested organization of these systems provides for basic reflexes, pain- and pleasure-related behaviours, drives and motivations, right up to the full flowering of emotion proper. He also acknowledges a connection to Spinoza's suggestion that drives, motivations and emotions are fundamental to our humanity.

Damasio develops this theme to make a powerful case that ethical behaviour has emerged from an overall programme of bioregulation. He asks what would be different in our behaviour without emotion and

feelings. We have partial answers to this question from detailed analysis of patients with damage to specific sectors of their brains. For example, patients with damage to the ventral prefrontal cortex have handicaps that manifest in amoral behaviour, a state that appears to reflect an inability to deploy social emotions such as shame, guilt, compassion and gratitude.

His suggestion that emotions and feelings are the bedrock of our ethical systems has the wider implication that an inborn capacity to acquire ethical norms depends on certain circuitry in the human brain. He suggests that a requirement for the display of ethical behaviour necessitates that categories of personal and social knowledge be connected with fundamental feelings, such as joy and sorrow.

A parallel theme to this extended account of human feeling is a search for the identity of the seventeenth-century philosopher. Spinoza stands out from his contemporaries by virtue of his resolute adherence to the rational, an attitude that makes him a supremely modern figure. He dared to deny the validity of the Bible's revealed truth, a stance that threatened every political structure that was underpinned by religious belief and dogma. He went so far as to argue that faith was "a mere compound of credulity and prejudices — aye prejudices, too, which degrade man from rational being to beast". Not surprisingly, such radical views made him an outcast from his family and community, culminating in a ritualistic excommunication from his synagogue.

Spinoza had a deep concern about the requirements for a good life, and devoted considerable attention to this in his *Theologico-Political Treatise*. Damasio contends that Spinoza's prescription for a life well lived — an ethical system and a democratic state — is of itself insufficient to achieve the sustained joy that Spinoza equated with human salvation. The added ingredient is a yearning borne out of an urgent need to give meaning to one's life. This yearning, he suggests, is a deep trait of the human mind that reflects a powerful biological mechanism for self-preservation: the conatus of Spinoza.

Looking for Spinoza is exceptionally engaging and profoundly gratifying. It achieves a unique combination of scientific exposition, historical discovery and deep personal statement regarding the human condition. It dares to ask how our accumulating knowledge of the human brain should inform the way we live our lives and organize our social world. Its erudition and wisdom provide a powerful statement that the pursuit of scientific knowledge about the human brain can go hand in hand with an overarching concern for our fellow humans. ■

Ray Dolan is in the Wellcome Department of Cognitive Neurology, University College London, Queen Square, London WC1N 3BG, UK.

Riding the solar wind

Journey From the Center of the Sun

by Jack B. Zirker

Princeton University Press: 2002. 312 pp.

\$29.95, £19.95

James A. Klimchuk

The Sun has been a source of awe and wonder since the dawn of humankind, but recent observational and theoretical advances have led to a surge of interest in our nearest star. Not unrelated is the growing recognition that outbursts from the Sun can have nasty effects on our increasingly technological society. Solar-induced 'space weather' causes power blackouts, disrupts communications, hampers navigation and surveillance, and damages or even destroys satellites. And this from the same temperamental object that sustains all life on Earth!

Journey From the Center of the Sun is a delightful book that surveys what we currently know about the Sun. It is primarily a popular-science book, accessible to anyone with an inclination for science and a background in high-school physics, but even professional astronomers will find it appealing. Jack Zirker presents the subject by taking us on a hypothetical journey. We begin in the ultra-hot core of the Sun, where energy is generated by nuclear fusion — enough energy in one second to power the United States for four million years. We then follow the flow of this energy outwards, first through the stable radiative zone and then through the roiling and churning convection zone. We eventually arrive at the solar 'surface', where the temperature is a relatively balmy 5,800 kelvin. After exploring for a while, we

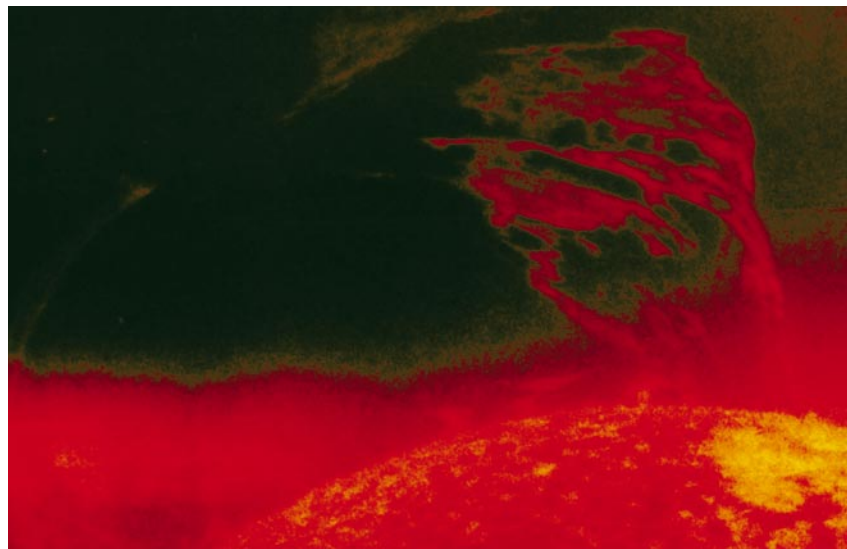
proceed upwards once again, into the tenuous but scorching corona. The temperatures here are measured in millions of degrees. Ultimately, we bid farewell to the Sun by hitching a ride on the outflowing solar wind.

Along the journey, we encounter a wide array of fantastic phenomena, including sunspots, magnetic loops, prominences, flares and coronal mass ejections. We learn that the average flare is equivalent to a billion megatons of TNT, and that Earth could fit comfortably inside a sunspot.

But the book is much more than solar phenomenology and gee whiz. Zirker explains the physical concepts behind the phenomena in a clear and insightful way, often making use of analogies from everyday life. Without equations, we learn about nuclear physics, thermodynamics and statistical mechanics, atomic physics and spectroscopy, and magnetohydrodynamics. This blending of physical explanation with observational description is the greatest strength of the book.

Also blended in are fascinating historical accounts of how many of the important discoveries came about, including insights into the personalities of the scientists involved. The story of how it was first realized that the corona is much hotter than the solar surface is especially interesting. What a shock it must have been to discover that the spectral line observation used to declare a new element, coronium, was actually caused by iron in a highly ionized state.

The book is not perfect, however. Zirker writes in a flowing, almost folksy style, and the book is highly readable for the most part, but some readers may find the later chapters a bit tedious. The text is heavily interspersed with useful figures, but those borrowed from technical papers are not always adequately explained, and the quality of image reproduction is disappointing. Picky experts may



Some like it hot: solar flares typically have the power of a billion megatons of TNT.

CORBIS

quibble about minor inaccuracies, and there are more than a few typographical errors.

But all in all, *Journey From the Center of the Sun* is a fine book. Solar physics is a vibrant field of research, with many space-based and ground-based observatories either operating or under construction, and many theoretical ideas being fleshed out with high-powered numerical simulations. Zirker has written an entertaining and educational survey of the current state of knowledge of our all-important neighbour that sits at the centre of the Solar System. ■

James A. Klimchuk is at the Naval Research Laboratory, Washington DC, USA.

Keeping up with evolution

Encyclopedia of Evolution

editor-in-chief Mark Pagel

Oxford University Press: 2002. 1,234 pp.

£185, \$325

Joel Peck

Evolutionary biology is the study of how populations of organisms change over time. Given the relatively small number of working evolutionary biologists, the field receives a surprisingly large amount of media attention. Indeed, evolution is one of the few areas of science that is of genuine interest to a substantial proportion of the public and to first-year undergraduates.

The field has grown enormously in recent years, in part because of progress in research on genomics, physiology, animal behaviour and other related topics. It is difficult to keep up with the key developments in evolutionary research, even for dedicated specialists, so the appearance of the *Encyclopedia of Evolution* is timely. This two-volume set comprises 365 articles from 330 different authors who cover all of the main areas of evolutionary biology and also explore many minor themes.

The encyclopedia opens with seven 'overview essays', which are written by some of the main contributors to modern evolutionary science, including John Maynard Smith, Jane Goodall and Luigi Luca Cavalli-Sforza. Some of these essays concern obvious organizing themes, such as the major transitions in evolution, macroevolution and the history of evolutionary thought. Others are instructive but rather idiosyncratic choices. The essays include pieces on motherhood, culture in chimpanzees, and darwinian medicine.

The rest of the encyclopedia is taken up with an alphabetical arrangement of short essays, which are generally well written and accessible. Each contains a bibliography, many of which include comments in the form of short annotations. These are very



A natural harbour

Do you recognize the harbour in this painting?

No, you probably don't, because it's not a painting at all, but the natural colouring on a section of polished paesina stone. This image, along with countless other artistic pictures of minerals and rock formations, can be found

in *Exposures of Mineral and Rock* by Dirk J. Wiersma (Inmerc bv, E49.50). This coffee-table book, which has text in both Dutch and English, is a stunningly illustrated guide to rocks, crystals, geodes, gems, fossils, minerals and all manner of things geological.

helpful for anyone trying to navigate an unfamiliar area of the literature. Equations are rare and technical terms are explained. Specialists should not expect to learn much in the way of facts about their own areas, but they may find new perspectives and insights. They will also find articles to bring them up to speed when they explore unfamiliar areas of evolutionary biology.

All of which is not to say that this encyclopedia is a perfect document. For one thing, it is heavy. The two volumes weigh in at more than 2 kg each, so you will not want to chuck them into your bag very often for a quick read on the bus. Also, I wish that some of the authors could have read drafts of each others' articles. For example, the article on altruism contains a discussion of group selection that doesn't seem to take into account recent advances in the theory underlying this phenomenon, and its list of further reading includes several articles from this encyclopedia, but not the excellent one on group selection!

Another problem is the selection of topics. Most of those that you would expect are here, but there are some important exceptions. There is no article on epistasis, despite the importance of this topic for many of the phenomena that are covered, such as speciation and the evolution of sexual reproduction. The evolution of genetic dominance is also a no-show. However, no encyclopedia can do everything. What is more interesting is the choice of some of the topics that have been included. There are two articles on art, for example, an article on warfare, and even one on globalization. Some of them have only

tenuous connections to evolutionary thinking. But more is probably better, and people come to evolutionary studies from many directions. Maybe some of the more far-out articles will help to direct unique and energetic minds towards evolutionary pursuits.

Many of the articles are written by researchers who have made key contributions to the areas about which they are writing. This is good in that the authors are close to the cutting edge of their topics and are able to convey the latest findings. The drawback is that, in some cases, the articles present an enthusiast's view of each topic, rather than the more even-handed account that might be expected had the editors employed professional science writers rather than researchers. Personally, I like the approach taken by the encyclopedia, but readers should recognize that they may have to read material from other sources to obtain a balanced view.

It seems almost certain that this encyclopedia will be a key source for researchers, as well as for students and teachers. Many students will turn to it for help when they start to write an essay or try to find a topic for research. They will find articles that provide a good initial orientation, along with the key references. Teachers will use it for preparing lectures and reading materials, but they will also need a copy to check for cheaters — this is likely to be one of the most widely plagiarized books in the history of evolutionary biology. ■

Joel Peck is at the Centre for the Study of Evolution, University of Sussex, Falmer, Brighton BN1 9QG, UK.

A voyage of discovery

Summarize yourself in the form of a title of a paper in Nature.

Thought experiments linking ocean ecology and biogeochemistry with species evolution.

What was your first experiment as a child (on pets/siblings/dead flies, and so on)?

When I was small, my elder sister suggested that I tease our sleeping dalmatian by poking him with a stick. He growled and we giggled. Suddenly he lost his temper, jumped up and bit me in the shoulder. The wound healed but the lessons remained.

Who has been the most important mentor in your career?

Johannes Krey, chair of the department of planktology at the Institute for Marine Research in Kiel. He took me under his wing on my arrival in Germany in 1964 at the age of 18. Krey was an exceptionally kind, sensitive man, of humorous inclination and excited by his science.

What makes a good scientific mentor?

This depends on the student. Some need to be guided to the details, others to the big picture. The ocean is still full of surprises so the good mentor maintains an open mind and teaches why honesty is the best policy — that is, a dispassionate interpretation of the full data set with particular attention to the outliers, which could be the exceptions that prove the rule.

What single scientific paper or talk changed your career path?

An article in the *Reader's Digest* entitled "Bread from the sea", which I read in my teens. I was always obsessed with biology so when I was offered a scholarship to study in Germany I picked marine biology and was sent to Kiel. There were famines in India at the time and I wanted to do something useful.

What book has been most influential in your scientific career?

Salim Ali's *Book of Indian Birds*, which captured my attention while I was still a toddler (according to my mother).

What gives you the most job satisfaction now? What are your major frustrations?

Discussing perplexing questions in marine science with enthusiastic, knowledgeable representatives from different disciplines; interpreting data on interdisciplinary research cruises with 50 scientists cramped in a small room on board *RV Polarstern* while weathering a storm in the Southern Ocean; creating structure and process in the

watery vastness around us from strings of numbers spewed from *in situ* sensors. What frustrates me most in science is territoriality. To counter the objection that interdisciplinary discussion is mere hand-waving, I coined the term 'heterodisciplinary fertilization'. Territorial males hate to be called homodisciplinary.

What literary character would you employ as a postdoc?

Puck. His magical powers would ensure success, he would be fun to work with, and I could quote Shakespeare in his reference letter.

What book currently resides on your bedside table?

Mind and Muscle by Elizabeth Langford.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Jim Corbett, who had much the same childhood as I did, in the Nainital district of Kumaon in the western Himalayan foothills, albeit 70 years earlier. He loved roaming the jungles, understood animals in their natural surroundings and matured into a passionate conservationist. He was a sincere, compassionate man who, at great personal risk and hardship, applied his knowledge (he called it 'jungle lore', I would call it non-verbal ecology) to track and shoot many cunning man-eating tigers and leopards that terrorized villages in remote areas. His stories (*Man-Eaters of Kumaon*, *The Temple Tiger* and so on) are grippingly authentic and beautiful to read. I wish I could have met this legendary, self-effacing hero.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

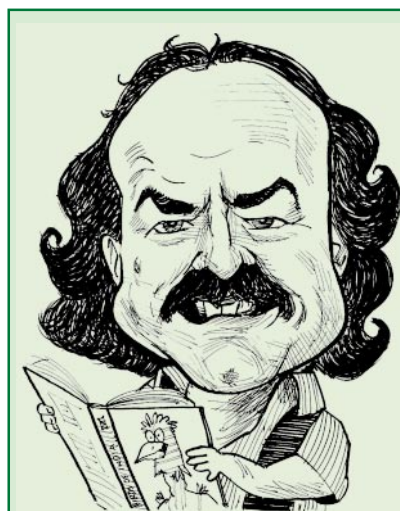
I would introduce myself and beg them to discuss any problems they might have encountered or perceived with my work.

What do you do to relax?

I relax in our garden and narrow strip of forest bordering the *Autobahn*. I stroll around, contemplate the landscape between bouts of hard work, keeping my back supple by squatting as much as possible. I also enjoy exploring the evolving mindset of our ancestors by active introspection: putting myself in their position, as down-to-earth as possible. Which is also why I like to squat and work.

What's the best piece of advice you've ever received?

When my father succumbed to my begging



Victor Smetacek

Victor Shaded Smetacek heads the Pelagic Ecosystems Division at the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven, Germany. He hails from India and loves natural and human history.

that I be allowed to go hunting with his shotgun at the age of ten, he made me memorize these lines of Victorian vintage: 'Never ever let your gun/ pointed be at anyone./ That it may unloaded be/ matters not the least to me.' He was worried about my safety and that of the many people who collected wood and fodder in the forest. His willingness to trust me made my sense of responsibility surge. I never even came close to an accident in the many subsequent years spent roaming the jungles.

What would you have become, if not a scientist?
I would have joined the Indian Forest Department and participated in conserving habitat and wildlife.

What single discovery, invention or innovation would most improve your life?

The discovery of the neural basis of the lust for power over others, and a treatment to correct this atavistic condition.

Name one extravagance you can now get away with because of your eminence.

Spending more time writing, which means going later to the office. I have always been a nightbird and write best under the partial exhaustion that only sets in in the wee hours of the morning.

What music would you have played at your funeral?

The hum of voices of friends meeting one another on the occasion of my passing on. ■

A blot on the land

William E. Rees

How many people lie awake at night pondering: "Just how much of the planet is dedicated to supporting me in the style to which I am accustomed?" This is — or should be — the first question of human ecology, yet very few people have given it a moment's thought. It seems that modern humans, particularly the growing proportion of urbanites, are committed dualists. People today are so psychologically alienated from nature that they rarely think of themselves as biological entities, let alone as dependent components of the world's ecosystems.

This is no trivial cognitive lapse. In 1992 the Union of Concerned Scientists warned that "a great change in our stewardship of the Earth and the life on it is required if vast human misery is to be avoided and our global home on this planet is not to be irretrievably mutilated". The world is facing an ecological crisis, yet most people have only the shallowest understanding of their own species' role as causative agent. True, everyone has something to say about climate change, biodiversity loss or some other so-called 'environmental problem', but the latter term effectively externalizes the issue (dualism again), drawing our attention from the cause to mere symptoms of the crisis. It is not just the 'environment' that needs to be fixed, but humans ourselves — the environmental crisis is the product of gross human ecological dysfunction (or, if you prefer, of humanity's spectacular evolutionary success).

One concept that seems to be effective in bringing this reality home is 'ecological footprint' analysis. Eco-footprinting actually answers the question with which this essay opens. It grabs attention because it focuses on personal consumption and translates it into a corresponding land area — something

else that ordinary citizens can understand.

I originally proposed ecological footprinting to counter economists' arguments that the concept of carrying capacity is irrelevant to human beings. Conventional carrying capacity is defined as the population of a given species that could be supported indefinitely by a specified habitat. However, economists argue that trade and improved resource productivity can raise the carrying capacity of a region indefinitely, thus rendering the concept meaningless for humans. Eco-footprint analysis subverts this argument simply by inverting the carrying-capacity ratio. Instead of asking how many people a particular area might support, we ask what area is required to support a particular population. Thus, the eco-footprint of a defined population is the area of productive ecosystems that is required on a continuous basis to produce the resources that the population consumes and to assimilate its waste. As the relevant land and water can be located anywhere, the method incorporates land and water areas appropriated through trade and natural flows, and automatically reflects prevailing technology.

The results of eco-footprint analysis are often startling, despite the fact that its basic method and assumptions are quite conservative. The *Living Planet Report 2002* shows that the average eco-footprints of residents of high-income countries range from almost five to over ten hectares, whereas the poorest of the poor live on less than half a hectare. Many high-income or densely populated countries have an ecological deficit that is several times larger than their domestic productive land/water areas. They survive on imports and by imposing on the global commons. Perhaps most disturbing is the fact that the world's average human eco-footprint is about 2.3 ha, even though there are only 1.9 ha of productive land and water per person on Earth. The human enterprise is already overshooting the global carrying capacity, funding further expansion by liquidating 'its' natural capital.

This situation is potentially catastrophic. Both the ecosphere and the human economy are self-organizing, dissipative structures that are far from (thermodynamic) equilibrium. This means that they grow and increase their internal order by importing available energy from their host systems and by dissipating their wastes back to their hosts. The problem is that whereas the ecosphere evolves by dissipating solar energy, the human enterprise grows by dissipating the ecosphere. Humans are consuming the ecosphere's wealth faster than the ecosphere can renew itself. We are now the dominant macro-consumer in all of the world's major ecosystem types, a paradoxical situation for a species that considers itself

Ecological footprints

Calculating the area of productive ecosystem required to support a population is a useful way to open people's eyes to the fact that we're stamping out the world's resources.

to have cut its ties with nature. The evidence is all around us: deforestation, species loss, landscape degradation, falling water tables and climate change.

What has brought us to this dangerous juncture? The driving forces are both biological and cultural, unconscious and overt. First, *Homo sapiens* is behaviourally predisposed to expand into all available ecological space and to consume to the level allowed by contemporary technology. Technology advances relentlessly, but evolution has provided humans with no inhibition against destroying their habitats. Second, humans remain a myth-bound species that is capable of astonishing feats of self-delusion. The dominant cultural myth today promotes a materialists' vision of global 'development', characterized by unlimited economic expansion and fuelled by open markets and more liberalized trade. This myth reinforces our already dangerous expansionist tendencies.

The general problem is apparently an ancient one. In 1995 Joseph Tainter wrote: "What is perhaps most intriguing in the evolution of human societies is the regularity with which the pattern of increasing complexity is interrupted by collapse..." Humanity is once again on a collision course with biophysical reality — this time on a global scale — and as biology provides no immediate remedy, the solution must be cultural. The eco-footprint challenge for both the natural and social sciences in the twenty-first century is to engineer the means by which human beings can live peaceful, comfortable and satisfying lives on the biological life-support provided by less than two hectares per capita (1.3 ha by 2050), while taking into account the needs of other species. The alternative is resource wars and descent into geopolitical chaos. ■

William E. Rees is at the School of Community and Regional Planning, University of British Columbia, 6333 Memorial Road, Vancouver, British Columbia V6T 1Z2, Canada.

SELECTED READINGS

- Tainter, J. *Futures* **27**, 397–407 (1995).
- Worldwide Fund for Nature *Living Planet Report 2002* (WWF, Gland, Switzerland, 2002).
- Rees, W. E. *Environ. Urbanization* **4**, 120–130 (1992).
- Rees, W. E. *Bull. Sci. Tech. Soc.* **22**, 249–268 (2002).

FR PHOTO CORBIS



Bigger is not better: people need to realize that their current consumption cannot be sustained.

The weight of expectation

C. D. Hoyle

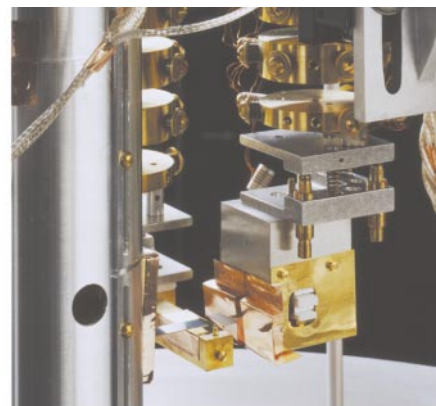
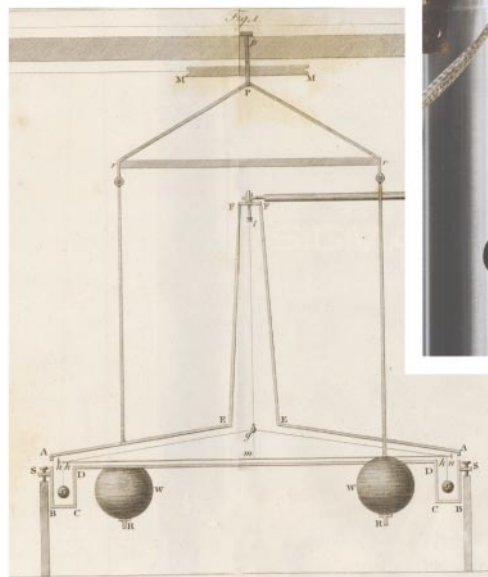
Newton devised his universal law of gravitation for planets, but does it work at small scales? A search for a deviation from the expected behaviour could provide the first evidence in support of string theory.

Gravity, one may think, is a rather well understood subject. It has been more than 300 years since Isaac Newton determined that the magnitude of the gravitational force, F , between two bodies of masses M and m (separated by astronomical distances) depends on the inverse-square of the distance, r , between them: $F = GMm/r^2$. And we have recently passed the two-century mark since Henry Cavendish made his first measurement¹ of the gravitational constant, G , which sets the fundamental strength of the interaction.

But developments in theoretical physics predicting possible deviations from the inverse-square law at small distances² have brought about a renewed enthusiasm for precise laboratory tests of this most familiar fundamental force. The results reported by Joshua Long and colleagues³ on page 922 of this issue describe an investigation of gravity at a previously unexplored scale. Furthermore, Long *et al.* have devised a new apparatus, in place of the low-frequency torsion balance that has been the workhorse of laboratory-scale gravitational experiments since the time of Cavendish^{4–6} — they use a kilohertz, resonant-oscillator technique to test the gravitational force experienced by test masses that are separated by just 108 μm (Fig. 1).

Gravity may have been the first of the fundamental forces to be described mathematically, but it is still the most poorly characterized. The inverse-square-distance dependence established by Newton had been assumed to hold true at short distances, but only recently has gravity actually been shown to exist between objects separated by less than one millimetre⁵.

Why is it so difficult to measure the properties of gravity, when it obviously exerts a strong influence on us every day? The reason is that gravity is incredibly weak compared with the other known fundamental forces of nature (electromagnetism, and the strong and weak forces). We feel gravity only because, unlike electromagnetism, it cannot be shielded with 'negative' masses — there is no gravitational repulsion to balance the attraction. To put the weakness of gravity in context, consider the classic example of the hydrogen atom: for the electromagnetic force between the proton and electron to be as small as the gravitational force they experience in the atom, they would have to



*Si Globorum duorum in se mutuo
gravitantium materia undique,
in regionibus quæ à centrīs equaliter
nubus quæ à centrīs equaliter distant,
homogenea sit: erit pondus Globi
alterutrius in alterum reciprocè ut
quadratum distantie inter centra.*

Figure 1 Gravity, past and present. In 1798, Henry Cavendish¹ used a torsion balance (left), which he had inherited from the Rev. John Michell, to determine the gravitational constant, G . Now Long *et al.*³ have devised a high-frequency resonator (inset) to explore the gravitational interaction on previously unexplored scales. Neither Cavendish nor Long *et al.* found any deviation from Newton's universal law of gravitation, set down in 1686 in his *Philosophiæ Naturalis Principia Mathematica*, and rendered here in its original Latin form.

be separated by about 2.5 million kilometres — roughly six times the distance between the Earth and the Moon.

Expressed another way, in any gravitational experiment the 'cancellation' of the electromagnetic interaction between test masses must be at the level of roughly one part in 10^{40} to leave any sensitivity to gravity — not an easy task on short distance scales, as local charge inhomogeneities and magnetic impurities in the materials of the experiment quickly become important. A careful analysis of subtle systematic effects is crucial for any such measurement.

Elegant experiments such as that of Long *et al.*³ have a broad appeal. Cosmologists and high-energy-physics theorists, for example, are interested in the short-range properties of gravity, and its theoretical aspects present at least as many difficulties as those encountered in the laboratory. Despite years of effort, it has not been possible to devise a quantum theory of gravity, even though most theoretical physicists believe that such a theory must exist to describe it in harmony

with the other three forces. Because gravity is so weak, its quantum behaviour is usually predicted to become important only at inaccessible distances, at the scale of the 'Planck length' (10^{-33} cm). It is disconcerting to many theorists that there is such a large discrepancy between the characteristic energy scale associated with this distance (10^{19} GeV) and that of the other three forces (10^3 GeV; $1 \text{ GeV} = 10^9$ electron volts). This is known as the 'hierarchy' or 'naturalness' problem.

Recently, string theories (so called because the fundamental particles are described as vibrations of one-dimensional strings)² have become the leading candidates for the successful integration of quantum mechanics and gravity. Such models require the existence of extra spatial dimensions, in addition to the three dimensions in which we live. In particular, one class of string-inspired theories⁷ attempts to solve the hierarchy problem by suggesting that gravity is just as strong as the other interactions, but that its strength is diluted because it propagates through many spatial dimensions; the

other fundamental forces are confined to our three-dimensional world, or 'brane'.

These extra dimensions would be curled up, or 'compactified', but could still be as large as a millimetre. If one were to look closely enough to begin to 'see' these large extra dimensions, gravity would behave as if it were propagating in a higher-dimensional world, and the inverse-square attraction would no longer be observed. In addition, different string-theoretical models predict as yet unseen fundamental particles that could mediate new interactions with strength comparable to that of gravity at sub-millimetre distances (the range of a possible new force would depend inversely on the particle's mass)⁸. Evidence of these new forces could also appear in short-range tests of the inverse-square law. So searching for exotic gravitational behaviour at the laboratory scale, instead of at the Planck length, could yield the first direct evidence for string theory.

The experiment by Long *et al.*³ is the latest search for deviations from the inverse-square law. Their results set the best constraints so far on any departure from that law between 10 and 100 μm , and have ruled out much — although not all — of the remaining parameter space for new forces. So far, Newton is holding his ground. Although this first-generation apparatus could not have detected any new effects that have the same strength as gravity, the

relatively high-frequency technique has advantages that may enable it eventually to probe shorter distances with higher sensitivity. For example, many sources of environmental noise, such as temperature and seismic fluctuations, have an intrinsically 'red' (or low-frequency) tendency; the high frequency of this experiment serves to suppress the effects of these disturbances.

The development of a promising technique to measure the unexplored properties of one of the fundamental forces in nature should always be welcomed, especially when the force in question is gravity. Having a 'gene pool' of experimental techniques involving scientists from many backgrounds contributes to the growth and refinement of the field as a whole. With advances in high- and low-frequency methods, we will be better able to understand and verify any surprises that gravity may hold. ■

C. D. Hoyle is in the Department of Physics, University of Trento, Via Sommarive 14, 38050 Povo (TN), Italy.

e-mail: hoyle@science.unitn.it

1. Cavendish, H. *Phil. Trans. R. Soc. Lond.* **17**, 469–526 (1798).
2. Pease, R. *Nature* **411**, 986–988 (2001).
3. Long, J. C. *et al.* *Nature* **421**, 922–925 (2003).
4. Lamoreaux, S. K. *Phys. Rev. Lett.* **78**, 5–8 (1997).
5. Hoyle, C. D. *et al.* *Phys. Rev. Lett.* **86**, 1418–1421 (2001).
6. Hoskins, J. K., Newman, R. D., Spero, R. & Schultz, J. *Phys. Rev. D* **32**, 3084–3095 (1985).
7. Arkani-Hamed, N., Dimopoulos, S. & Dvali, G. *Phys. Lett. B* **429**, 263–272 (1998).
8. Kaplan, D. B. & Wise, M. B. *J. High Energy Phys.* **8**, 37 (2000).

Ecology

The how and why of biodiversity

Kevin J. Gaston

A study of reef fish in the Indian and Pacific oceans reveals that the structures of local communities and their regional context are intricately entwined. New species spread far from an oceanic 'hotspot' of diversity.



Figure 1 Cradle of diversity: the islands (centre) of the Indonesian and Philippine region.

In the field of biodiversity, Rudyard Kipling's "six honest serving men"¹ have progressed at very different speeds. What, Who, Where and When have established much about the basic distribution patterns of life across the Earth. But How and Why have been struggling, and their difficulties have been well aired^{2,3}. The problem has not been an absence of ideas — if anything, quite the reverse. In particular, innumerable theories have been proffered to explain why there are more species in some areas than in others. But rigorous testing and discrimination among a core group of possible mechanisms has proved difficult. On page 933 of this issue⁴, Mora and colleagues give How and Why a helping hand.

The mechanisms that might drive patterns of species richness have traditionally been divided between abiotic variables, such as temperature, rainfall and geology, and biotic variables, such as competition, preda-

tion and parasitism. But it has become increasingly apparent that an equally, if not more, critical issue may be the distinction between processes that operate at different spatial scales. Ecologists have been most comfortable considering mechanisms that operate at spatial scales similar to those at which their own senses principally function — from metres to tens or perhaps a few hundreds of metres. So most ecological studies have been, and continue to be, conducted in highly restricted areas; most experiments have changed little in unit size (although they are now almost invariably replicated) since Darwin⁵ established his 'worm stone' to determine the rates at which earthworms bury material. Although many important insights have followed, a general understanding of why local species assemblages and communities are structured in the way they are, without recourse to case-by-case contingencies, has largely remained elusive⁶.

One obvious possibility, recognized long ago, is that processes operating over much broader spatial scales — at which experiments are difficult or impossible to conduct — markedly dictate what is seen locally. A local community is assembled from a regional pool of species. The size and structure of the pool are influenced by regional processes, including the geophysical properties and history of the region (its age, geology, size and climate), and broad-scale ecological or evolutionary processes such as species migrations and invasions, speciation and regional extinction^{6,7}. These determine features of the pool, for example the total number of species, the number of individuals within each species, the size of the individuals, and their feeding groups.

Mora *et al.*⁴ now elegantly demonstrate that the structure of a local species assemblage and its regional context are indeed inseparable. Basing their analyses on almost 2,000 species of reef fish across 70 locations in the Indian and Pacific oceans, they describe three important findings.

First, most of the species that occur at these sites are also present at a centre of high diversity in the Indonesian and Philippine region (IPR); the existence of this centre is well established. Second, the structure of the local assemblages is strongly influenced by their distance from the IPR, such that as this distance increases — both latitudinally and longitudinally — so the number of species declines. Third, as the distance from the IPR increases, the likely dispersal ability of those species present also increases; sites farther from the IPR tend to comprise species with longer open-water larval stages. In short, the local assemblages of reef fish at sites in the Indian and Pacific oceans are in large part shaped by which species have been able to reach them from the IPR. On a regional scale, the patterns of species richness have likewise been shaped by how many

species have reached different localities.

Hotspots of species richness have variously been argued as representing cradles of diversity exhibiting high origination rates, museums of diversity exhibiting low extinction rates, or some combination of the two. For reef fish, the IPR seems to function as the first of these, constituting a major (but not exclusive) centre of speciation (new species formation) in the Indian and Pacific oceans, from which species disperse to more peripheral locations. Other major marine centres of origin, functioning for other organisms, seem more obviously to have served as museums, as well as cradles, of diversity⁸.

The high levels of speciation in the IPR may result from the high number and density of islands (and hence reefs; Fig. 1). This would theoretically enhance allopatric speciation (the differentiation of geographically isolated populations into distinct species) and perhaps other forms of speciation, resulting in the IPR also being a hotspot of endemism.

In addition, a negative relationship has recently been demonstrated, both theoretically and empirically, between the size of a species' geographical range and its probability of speciation^{9,10}. This finding raises the intriguing possibility that hotspots of origination experience positive feedbacks, in which to some degree the restricted species ranges that result from high levels of speciation encourage more speciation. This is because species with small ranges are characterized by reduced dispersal between different populations and by lower local densities, which reduce levels of gene flow⁹ and hence enhance the likelihood of genetic isolation.

Positive correlations between range size and dispersal ability have been reported for other groups of marine organisms, and would fit with the findings of Mora *et al.*⁴ for reef fish in the Indian and Pacific oceans. Positive correlations between range size and local abundance are very general, and might also apply to reef fish: although there is little evidence of this at present, insufficient data are available for most families¹¹. So the very reasons for the existence of the centre of diversity in the IPR may also be responsible for its importance in the structuring of reef assemblages across the Indian and Pacific oceans.

Kipling sent his honest serving men "east and west" and "over land and sea". In seeking to understand patterns of species richness, such a broad-scale view was undoubtedly the correct one. ■

Kevin J. Gaston is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

e-mail: k.j.gaston@sheffield.ac.uk

1. Kipling, R. *Just so Stories for Little Children* (Macmillan, London, 1902).
2. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
3. Gaston, K. J. *Nature* **405**, 220–227 (2000).
4. Mora, C., Chittaro, P. M., Sale, P. F., Kritzer, J. P. & Ludsins, S. A. *Nature* **421**, 933–936 (2003).
5. Darwin, C. *The Formation of Vegetable Mould through the Action of Worms* (Murray, London, 1888).
6. Lawton, J. H. *Community Ecology in a Changing World* (Ecol. Inst., Oldendorf/Luhe, 2000).
7. Huston, M. A. *Oikos* **86**, 393–401 (1999).
8. Briggs, J. C. J. *Biogeogr.* **30**, 1–18 (2003).
9. Gavrillets, S., Li, H. & Vose, M. D. *Evolution* **54**, 1126–1134 (2000).
10. Jablonski, D. & Roy, K. *Proc. R. Soc. Lond. B* **270**, 401–406 (2003).
11. Jones, G. P., Caley, M. J. & Munday, P. L. in *Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem* (ed. Sale, P. F.) 81–101 (Academic, Amsterdam, 2002).

Geochemistry

Lost terrains of early Earth

Stein B. Jacobsen

Isotope data provide insight into the earliest phases of terrestrial evolution. The latest reappraisal supports the view that the early Earth had a cratered crust which crystallized from a magma ocean.

Unlike Earth, several bodies such as the Moon, Mercury and Mars have a clear record of their surface history dating back almost to the origin of the Solar System. This early crust is now no longer evident on Earth, where the oldest rocks are 3.8–4.0 billion years old and thus date to about 600–700 million years after the planet's formation. Studies^{1–3} in which neodymium (Nd) isotopes were used as a window into these lost terrains concluded that the Earth probably had a magma ocean from which a proto-crust and a layered mantle crystallized about 4.47 billion years ago. But the interpretation was challenged by reports^{4–7} that there was no evidence of an expected corollary of that conclusion — a specific signature of another isotope,

hafnium (Hf). In particular, it was proposed^{6,7} that before 4.0 billion years ago the crust was similar to the continental crust we see today.

On page 931 of this issue, however, we have a reversal of perspective. In a new study of Hf isotopes in meteorites, Bizzarro *et al.*⁸ point out that the data used up to now for a crucial parameter, the evolution of the lutetium (Lu)–Hf isotopic system in 'bulk' planetary material^{9,10}, were incorrect. With their new data, the Hf and Nd information fall into line, thus supporting the earlier inferences drawn from the Nd isotopes. Appropriately enough, the discovery of Hf was announced in these pages 80 years ago, almost to the month (see Box 1).

Chondritic meteorites — aggregates of



100 YEARS AGO

A few weeks ago we described some of the excellent results obtained by Messrs. Heraeus, of Hanau, in their attempts to produce apparatus of "silica glass," and Prof. Dewar had added point to our remarks by exhibiting at the Royal Institution a "liquid air holder" made of silica, which had been made to order and sent by return of post, almost, from Hanau to London a few days before. Similar apparatus could have been made in England, it is true, but it could not have been produced by any means so quickly as at Hanau... Truly, as Prof. Dewar said the other evening, there will soon be another "lost industry" if our practical men do not wake up. Silica glass making as an industry no doubt is still in earliest infancy, but though so young, it already shows signs of growth. Everyone who has worked with silica, and knows its properties and how comparatively easy it is to work with, foresees that soon silica glass will replace ordinary glass in many of its most important applications.
From *Nature* 26 February 1903.

50 YEARS AGO

G. Notini and S. Forselius discuss the methods which have been undertaken to exterminate the wild rabbits on Gotland Island. The wild rabbit was introduced into Sweden with the object of providing a new game animal of commercial value. Vigorous stock was selected and care was taken to ensure the proper environmental conditions, based on European accumulated experience in parts where the stock had become more or less stabilized, excess numbers being kept down by small predatory animals and also disease. As has occurred in other parts of the world where mammals, birds and plants have been introduced outside their own habitat, the rabbits in Gotland increased rapidly in numbers. None of their ordinary checks was present, the only one being the occasional severe winters experienced in the island. The ordinary methods of man-shooting, poisoning, snaring, etc. — but not poison, have proved ineffective; poison is regarded as too dangerous... The rabbit to-day in Gotland constitutes such a menace to silviculture and agriculture that it is ranked with the small rodents. Work is now being undertaken on the introduction of the virus disease *Myxomatosis cuniculi* into the Gotland rabbit population.
From *Nature* 28 February 1953.

Box 1 Hafnium history

Not only is 2003 the eightieth anniversary of the identification of hafnium but, like the paper by Bizzarro *et al.*⁸, news of the discovery came from Copenhagen — hence the element's name, which derives from *Hafniae*, the Latin for Copenhagen. The paper appeared in *Nature* on 20 January 1923 (111, 79; 1923) under the title "On the missing element of atomic number 72".

The discovery was made by Dirk Coster (a Dutchman) and George de Hevesy (a

Hungarian) working in Niels Bohr's laboratory. Initially, the nature of element 72 was a little controversial — some claimed that it had been found among the rare earth metals. But Bohr believed that it should lie not there but among the group 4 elements with zirconium; and so it turned out. Following X-ray spectroscopy of zirconium minerals, Coster and de Hevesy concluded that: "It seems to be very probable that ordinary zirconium contains at least from 0.01

to 0.1 per cent. of the new element. Especially the latter circumstance proves that the element 72 is chemically homologous to zirconium."

De Hevesy had a peripatetic career, working not only in Copenhagen, but in Manchester (with Rutherford), Vienna, Budapest, Freiburg, Cornell and Ghent. In 1943, he won the Nobel Prize in Chemistry for his research on the use of isotopes as tracers for studying chemical processes.

Tim Lincoln

material from the primitive Solar System that have not undergone planetary differentiation — are central to studies such as these. They provide the best estimate of the starting composition of the terrestrial planets except for the most volatile elements (such as hydrogen, helium and nitrogen). Therefore it is commonly accepted that the average evolution of the chondritic Lu–Hf and samarium (Sm)–Nd systems also represent their evolution in the terrestrial planets as a whole. Deviation from the chondritic evolution of these isotopic systems is thus a measure of planetary differentiation processes such as mantle melting and solidification of these melts into crustal or mantle layers.

The basis for modelling planetary evolution with isotope systems such as Lu–Hf or Sm–Nd is a good estimate of the bulk planetary isotope evolution of such a system. These parameters have long been well established for the Sm–Nd system^{11,12}. But because of the difficulty of making precise Lu/Hf isotopic measurements on meteorites, the parameters used for this system^{9,10} were of inferior quality. Bizzarro and colleagues' study⁸ of chondritic meteorites rectifies this situation and is also in agreement with a Lu/Hf study of a group of meteorites called eucrites¹³ (eucrites are thought to be the product of lava flows from the surface of the asteroid Vesta, formed as partial melts of its interior).

The decay of ¹⁷⁶Lu (half-life 36 billion years) into ¹⁷⁶Hf is a good clock for tracing the evolution of planetary crusts and mantles, because the behaviour of these elements during melting and crystallization is well understood and produces variations in concentration that can be quantitatively predicted. The process gives a crust and mantle of different chemical compositions. Compared with the primordial, well-mixed Earth, the composition of which is also known from the analysis of meteorites, the crust has a higher proportion of some elements than the mantle. Lutetium and Hf are

preferentially extracted from the mantle and enriched in the crust, Hf more so than Lu. The Sm–Nd isotopic system has an entirely analogous behaviour.

The Moon has a distinctive signature for the Nd and Hf systems, and it is well accepted that it was produced by a particular evolutionary pattern — an early magma ocean gave way to crust overlying accumulated mineral layers of the lunar mantle, the main differentiation occurring around 4.4 billion years ago. Similar events had been inferred from Nd isotopes for the early Earth. But this view was challenged by studies of Hf isotopes, on the grounds that the Nd data had been altered by metamorphism whereas the Hf data were more robust and gave the true signature of early differentiation.

Bizzarro *et al.*⁸ have now found a clear Hf isotope record of the Earth's primordial crust. From Hf isotope data of 3.1–4.1-

billion-year-old zircons — resistant mineral grains that have survived from the time of the parent rock's formation, even when that rock has not — they conclude that the early differentiation between crust and mantle occurred 4.33 ± 0.10 billion years ago (about the time of early differentiation of the Moon). Although this early terrestrial crust no longer exists, it left a clue behind in that the zircon Hf isotope ratios are much higher than the bulk planetary Lu–Hf isotopic evolution as inferred from meteorites — that is, much higher than in the primordial mantle. So the host rocks of the ancient zircons did not emerge from mantle that had never given birth to crustal rocks before. Rather, the mantle source of the rocks containing the zircons had already supplied some Hf to an even earlier crust. This is in line with what has been observed for two Sm/Nd isotopic clocks: the decay of ¹⁴⁷Sm (half-life of 106 billion years) to ¹⁴³Nd, and of ¹⁴⁶Sm (half-life of 103 million years) to ¹⁴²Nd.

Today, Earth's crust forms mainly at mid-ocean ridges where crustal spreading occurs, and is recycled into the mantle where tectonic plates collide. The Earth's early crust was probably not formed by this mechanism. If plate tectonics had been active then, the isotopic effect observed in Hf and Nd isotopes would not have survived. So the early crust must have been relatively stable, and any recycling to the mantle must have started relatively late. The most likely cause of the total destruction of Earth's early crust is the late meteorite bombardment in the inner Solar System, which occurred about 3.9 billion years ago and is clearly preserved in the lunar crustal record as the 'terminal lunar cataclysm'¹⁴. Other planetary crusts from this time are heavily cratered (as shown for the Moon in Fig. 1), and this must also



Figure 1 Lunar analogy. These heavily cratered highlands on the Moon show what the crust on the early Earth might have looked like. Later terrestrial processes, including plate-tectonic movement, have destroyed any direct information of this initial phase of Earth evolution.

NASA

have been true of the early Earth. This crust was later destroyed by further impacts and possibly by the onset of the plate-tectonic cycle some 3–4 billion years ago.

The early age of the proto-crust, as determined with Hf isotopes, requires that the now-extinct ^{146}Sm was common and enriched in the most ancient rocks compared with chondritic meteorites, and should give a signature of a high ^{142}Nd abundance. To add to the descriptions of ^{142}Nd anomalies in rocks from the 3.8-billion-year-old rocks in west Greenland^{2,3}, there are now reports that such anomalies are relatively common in samples from this area^{15,16}. From the size of the ^{142}Nd anomaly, one can infer that Earth's most ancient crust must have started forming before Earth was 100 million years old — that is, 4.47 billion years ago.

Bringing the Nd and Hf data into agreement is a great step forward. With that done, we can expect further, exciting results about the earliest crust on Earth to be forthcoming. ■

Stein B. Jacobsen is in the Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA. e-mail: jacobsen@neodymium.harvard.edu

1. Jacobsen, S. B. & Dymek, R. F. *J. Geophys. Res.* **93**, 338–354 (1988).
2. Harper, C. L. Jr & Jacobsen, S. B. *Nature* **360**, 728–732 (1992).
3. Jacobsen, S. B. & Harper, C. L. Jr in *Earth Processes: Reading the Isotopic Code* (eds Basu, A. & Hart, S.) 47–74 (American Geophysical Union, Washington DC, 1996).
4. Vervoort, J. D. *et al.* *Nature* **379**, 624–627 (1996).
5. Vervoort, J. D. & Blichert-Toft, J. *Geochim. Cosmochim. Acta* **63**, 533–556 (1999).
6. Amelin, Y. *et al.* *Nature* **399**, 252–255 (1999).
7. Amelin, Y., Lee, D.-C. & Halliday, A. N. *Geochim. Cosmochim. Acta* **64**, 4205–4225 (2000).
8. Bizzarro, M., Baker, J. A., Haack, H., Ulfbeck, D. & Rosing, M. *Nature* **421**, 931–933 (2003).
9. Patchett, P. J. & Tatsumoto, M. *Nature* **288**, 571–574 (1980).
10. Blichert-Toft, J. & Albarède, F. *Earth Planet. Sci. Lett.* **148**, 243–258 (1997).
11. Jacobsen, S. B. & Wasserburg, G. J. *Earth Planet. Sci. Lett.* **50**, 139–155 (1980).
12. Jacobsen, S. B. & Wasserburg, G. J. *Earth Planet. Sci. Lett.* **67**, 137–150 (1984).
13. Blichert-Toft, J. *et al.* *Earth Planet. Sci. Lett.* **204**, 167–181 (2002).
14. Tera, F. *et al.* *Earth Planet. Sci. Lett.* **22**, 1–21 (1974).
15. Boyet, M. *et al.* *Geochim. Cosmochim. Acta* **66**, A99 (2002).
16. Caro, G. *et al.* *Geochim. Cosmochim. Acta* **66**, A120 (2002).

Developmental biology

A twist in a mouse tale

Nick Dyson

Studies of the retinoblastoma gene can still deliver surprises, and enlightenment. Several of the abnormalities in mice lacking this gene are, it seems, the indirect consequence of a placental defect.

On page 942 of this issue, Wu *et al.*¹ describe a surprising new observation about the function of *Rb*, the gene that is mutated in retinoblastoma cancers. *Rb* is often described as a prototype for tumour suppressors — that priceless category of genes that, among other tasks, protect us from developing cancers — as it was the first gene of this type to be identified. The biochemical and cellular properties of its protein product have been investigated in great detail, and the characteristics of *Rb*-deficient mice have been studied for over a decade. One might think that there is little left to be learned. But how wrong that would be! The results reported by Wu *et al.* force us to view the abnormal development of *Rb*-deficient mice from a completely new perspective.

Identifying the function of a gene is a tough challenge. A classic strategy is to inactivate the gene in model organisms such as mice, then sit back and see what happens. This approach has particular value when applied to tumour suppressors. Many genes that have an effect on the proliferation of cancer cells also influence normal cellular behaviour during animal development — and the tumour suppressors are no exception. Changes in the physical and behavioural characteristics (the phenotype) of an

organism after the inactivation of a tumour suppressor show where the gene is essential, and give insights into events that are important for animal development. The mutant phenotypes also provide clues to the types of biological processes that might be affected by loss of the tumour suppressor during cancer development.

Mice that lack both copies of the *Rb* gene are highly abnormal, and die between days 14 and 15 of embryonic development^{2–4}. The embryos show defects in the cell-division cycle and in cell differentiation, as well as high levels of cell death. Severely affected tissues include the central nervous system and liver. *Rb*-deficient embryos also have defects in red-blood-cell development, and it was originally suggested that anaemia was a likely cause of death.

Experiments with cell culture have further revealed that the lack of Rb protein can compromise the normal control of cell-cycle exit, cell differentiation and cell survival, as well as the control of 'checkpoints' by which the cell monitors the fidelity of cell-cycle progression. Because the Rb protein is broadly expressed and is normally present in the tissues that are affected in *Rb*-deficient animals, many have assumed that the phenotypes of these animals were caused simply by

the loss of Rb protein from the affected tissues.

Wu *et al.*¹ now characterize a placental abnormality in *Rb*-mutant mice that was overlooked in earlier studies. Placental architecture depends on cells derived from both mother and developing embryo. Wu *et al.* find that the defect in *Rb*-deficient placentas seems to be caused by the overproliferation of extra-embryonic trophoblast stem cells. This affects the labyrinth layer — a layer in the mouse placenta that contains the fetal-maternal interface and is morphologically different from the human placenta⁵ — and results in an over-representation of trophoblasts and a decrease in blood spaces. The consequence, as measured by the accumulation of essential fatty acids in normal and mutant embryos, is decreased nutrient transport from mother to embryo across the placenta.

This finding might have been dismissed as simply another tissue that requires the Rb protein. But Wu *et al.* take our understanding of *Rb*-mutant mice to a higher level, by showing the significance of the defect that they have found. They use two different genetic tools that make it possible to generate an *Rb*-deficient embryo supplied with a normal placenta containing wild-type extra-embryonic cells. Remarkably, embryos prepared by both methods develop to term, and many of the defects previously described for *Rb*-deficient embryos (including the defects in the central nervous system and in red-blood-cell differentiation) are suppressed. So, the placental abnormalities seem to be the primary reason why the *Rb*-mutant mice die at such an early embryonic stage. Several of the defects that were thought to reflect the loss of Rb function in specific tissues are instead indirect effects of a more systemic change.

Interestingly, however, several abnormalities still occur. These include defects in muscle and bone, cataracts, and inappropriate cell proliferation in several tissues. These residual abnormalities might reflect a specific requirement for the Rb protein in each of these tissues, but further experiments are needed to be sure.

This result sheds light on a long-standing puzzle. Several years ago the analysis of chimaeric mice, composed of random mixtures of wild-type and *Rb*-deficient cells, led to the surprising observation that the mutant cells can make a substantial contribution to a wide assortment of apparently normal tissues^{6,7}. This showed that many of the abnormal features of *Rb*-deficient embryos could be prevented by the presence of wild-type cells — but the reason for this effect was unclear. The placental defect described by Wu *et al.*, and its consequences, might finally explain why animals composed entirely of *Rb*-deficient (embryonic and extra-embryonic) cells are so much more severely affected than mosaic animals.

The findings also have an impact on studies of compound mutants in which the

effects of mutating *Rb* are partly overcome by mutating a second gene (*E2f1*, *E2f3* or *Id2*) that encodes an Rb-binding protein^{8–10}. These genetic interactions were interpreted by assuming that Rb and the Rb-binding proteins interact functionally in the tissues that are affected in *Rb*-deficient animals. The results of Wu *et al.* raise a whole new set of questions. For example, mutation of *E2f1* or *E2f3* in *Rb*-deficient animals prevents the cell death that occurs in the central nervous system of *Rb*-deficient mice^{8,9}. So, do the products of the *E2f* genes have a specific role in cell death in the nervous system? Or might a reduction in *E2f* activity alleviate the placental defect of *Rb*-deficient embryos and indirectly protect the central nervous system?

Inevitably, there are further questions. For instance, is an *Rb*-mutant placenta enough to cause embryonic death, as well as defects in the central nervous system and red-blood-cell development, in wild-type embryos? Perhaps the homozygous mutant phenotypes result from a combination of factors, requiring both the systemic changes caused by the placental defects and also intrinsic changes resulting from the loss of Rb protein in the affected tissues.

And what are the properties of trophoblast stem cells that make them particularly sensitive to the loss of *Rb*? Despite the tremendous progress in molecular studies of the Rb protein in the 16 years since its gene was identified, it remains uncertain which of the many cellular and biochemical functions attributed

to it are important for its role as a tumour suppressor. Retinoblastoma, a cancer of the retina, occurs only in very young children, and this temporal boundary is thought to reflect a window in eye development where a poorly defined group of cells is especially sensitive to the loss of *Rb*. One of the challenges is to work out which cell types need the Rb protein and why it becomes so important at certain times in development. Does its function vary, or does the 'wiring' of cells change so that Rb becomes critical? Further studies of *Rb*-deficient animals might provide clues.

Mutant phenotypes often result from complex interactions between multiple cell types, and one of the attractions of animal studies is the opportunity to study and dissect these effects. The work of Wu *et al.*¹ is a beautiful illustration of how even the best-studied animal phenotypes can have unexpected components.

Nick Dyson is at the Massachusetts General Hospital Cancer Center, Building 149, 13th Street, Charlestown, Massachusetts 02129, USA.
e-mail: dyson@helix.mgh.harvard.edu

1. Wu, L. *et al.* *Nature* **421**, 942–947 (2003).
2. Lee, E. Y.-H. P. *et al.* *Nature* **359**, 288–294 (1992).
3. Jacks, T. *et al.* *Nature* **359**, 295–300 (1992).
4. Clarke, A. R. *et al.* *Nature* **359**, 328–330 (1992).
5. Rossant, J. & Cross, J. C. *Nature Rev. Genet.* **2**, 538–548 (2001).
6. Williams, B. O. *et al.* *EMBO J.* **13**, 4251–4259 (1994).
7. Maandag, E. C. *et al.* *EMBO J.* **13**, 4260–4268 (1994).
8. Tsai, K. Y. *et al.* *Mol. Cell* **2**, 293–304 (1998).
9. Ziebold, U., Reza, T., Caron, A. & Lees, J. A. *Genes Dev.* **15**, 386–391 (2001).
10. Lasorella, A., Nosedà, M., Beyna, M., Yokota, Y. & Iavarone, A. *Nature* **407**, 592–598 (2000).

Oceanography

The brawniest retroflexion

Arnold L. Gordon

The influence of the Agulhas system of currents and eddies around southern Africa extends far beyond that region. Hence the especial need for a better understanding of the complex phenomena involved.

An intriguing class of ocean-circulation pattern is that exhibiting a retroflexion — one in which a swift current, flowing along the western boundary of a continent, separates from the continental margin and curls back upon itself. Large buoyant pools of water detach from the resulting loop, injecting eddies into a neighbouring ocean province. The ocean's most pronounced retroflexions are associated with the Agulhas, Brazil, North Brazil and East Australian currents, and the Gulf of Mexico's Loop Current. The most energetic of these is the Agulhas retroflexion, for which the term was designated¹. This brawniest of retroflexions is the subject of a special issue of a journal² — *Deep-Sea Research II* — in which 12 papers describe both observations and simulations of the retroflexion, and the consequential

exchange of water between the Indian and Atlantic oceans.

The geography of the Agulhas system is shown in Fig. 1. On rounding the southern rim of Africa, the Agulhas Current abruptly turns anticlockwise, to flow back to the Indian Ocean as the Agulhas return current. The special nature of the Agulhas retroflexion stems from the unique regional geography and wind patterns. The southern coast of Africa is about 5° of latitude closer to the Equator than the westerly wind maximum, the latitude where western boundary currents in the subtropics are expected to separate from the continental margin to turn into their ocean's interior. The Agulhas 'runs out' of western continental margin before the wind allows for such separation. The resulting retroflexion is shaped by the sea-floor morphology and by regional to large-scale winds³; its form is expected to

alter with changes of the maximum westerlies, or with the strength of the Agulhas Current. The warm surface water trapped within the Agulhas retroflexion transfers heat to the atmosphere, the largest such exchange in the Southern Hemisphere. The effects are considerable, for instance in influencing rainfall patterns as far away as Australia⁴ or deep-ocean 'overturning' in the northernmost Atlantic⁵.

The significance of the Agulhas retroflexion does not end with its momentary loop into the southeast corner of the South Atlantic, for there are its outputs and inputs to consider. A major output, for instance, is the considerable 'leakage' of water from the Indian Ocean into the upper kilometre of the Atlantic Ocean. The magnitude of this phenomenon has been the subject of conflicting results and much debate over the past 15 years, as have the wider implications^{6,7}. These include the effects of Agulhas leakage on the heat flow towards the Equator within the South Atlantic, and on the formation of North Atlantic Deep Water (NADW). Deep-water formation in the North Atlantic is one of the central features of ocean circulation. This overturning circulation involves water movement on a massive scale — the surface flow of warm water northwards, its sinking at high latitudes, and the return flow of cooler water at depth.

The various and widespread inputs to the Agulhas system are shown in Fig. 1. Streams of water are derived from far-distant parts of the Pacific Ocean, and may be part of a NADW-induced global balance^{8–11}. Red Sea intermediate water is part of the picture^{12,13}, as is eastward-flowing water in the South Atlantic, both of which contribute to the mix in the Indian Ocean. All of the return pathways must pass through or become entangled with the Agulhas retroflexion, which may act as a 'valve' regulating the buoyancy of water in the upper kilometre of the South Atlantic Ocean and may in turn regulate NADW overturning.

Clearly, then, the Agulhas retroflexion presents a challenge to oceanographers. Equally clearly, the convenient but unrealistic notion of a 'steady-state ocean' is particularly deficient in portraying this system. This is where the *Deep-Sea Research II* special issue² comes in. Comprehensive time-series observations (*in situ* and from satellites) and accurate computer simulations are required to capture the detailed behaviour of the retroflexion and its effects. The various contributors describe the advances that have been made on both fronts.

Much information has been provided by acoustically tracked instruments, known as RAFOS — Ranging and Fixing of Sound — floats, that were allowed to drift at depths ranging from 200 to 1,350 m. The data show that the mean circulation in the Cape Basin (the first Atlantic basin to receive leakage of Agulhas water) is obscured by a cauldron of

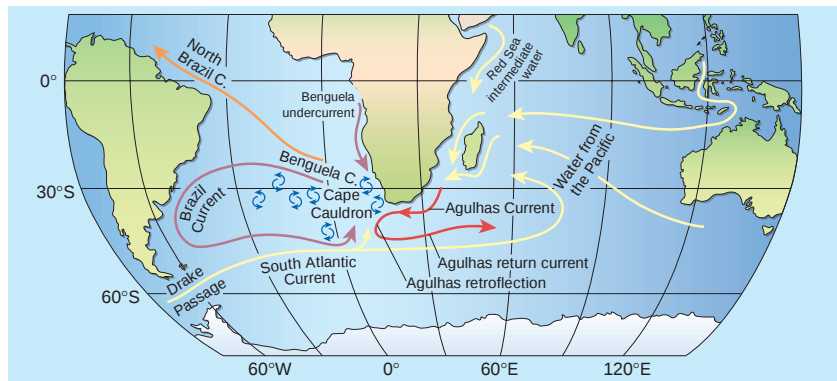


Figure 1 The Agulhas system and associated flow patterns. The Agulhas Current draws water from the Pacific Ocean through the Indonesian throughflow and Drake Passage, and from the Tasman Sea. It abruptly turns back towards the Indian Ocean near 20° E. Here, at the Agulhas retroflection, 'leakage' of water occurs within an array of cyclonic (clockwise) and anticyclonic (anticlockwise) eddies that are injected into the vigorous stirring and mixing environment of the Cape Basin (the 'Cape Cauldron'). The South Atlantic Current adds water to the cauldron, as does the subsurface flow of the Benguela undercurrent along the west coast of Africa, and salty Red Sea water along the east coast. A blend of these waters spreads into the Brazil and North Brazil currents. The latter is part of the large-scale 'overturning circulation' induced by the formation of North Atlantic Deep Water.

powerful eddies, 100–400 km in diameter. All of these eddies are generated within or along the periphery of the retroflection, some of them originating from the Indian Ocean, others from the South Atlantic Current (Fig. 1). These eddies are not just large anticyclonic rings, cast off from the retroflection at an average rate of six per year. Rather, they are a hodgepodge of types, which interact with each other and with the main retroflection, in a general milieu of vigorous stirring and mixing. Indian Ocean water is trapped within the eddy cores, and is often lost in the process, blending into the regional background (this is usually 'thermocline water' — that is, water from above about 800-m depth — but can take in the upper kilometre or two).

Studies of specific, newly formed Agulhas rings have exposed their turbulent birth and early evolution as they drift into the cauldron, with numerical simulations adding further detail to our picture of the complex Cape Basin circulation. These simulations can resolve quite fine-grained behaviour, and they reveal the coexistence, in dipoles, of anticyclonic eddies intensified at the surface with cyclonic partners intensified at the sea floor. The cyclonic eddies are caged in by the topography of the Cape Basin. But the anticyclonic eddies can break out of the basin and enter the South Atlantic, although with substantial energy loss.

Other authors in the special issue² describe how the witch's brew of the Cape Cauldron feeds blended water, via the Benguela Current, into the subtropical gyre of the South Atlantic. Some of this water remains within the gyre, eventually finding its way back to the Indian Ocean in what can be thought of as a super-subtropical gyre spanning the South Atlantic and Indian Oceans. But some enters the Northern Hemisphere within the North

Brazil Current, as part of the NADW overturning circulation. About three Agulhas rings shed from the retroflection survive the blender, and make stately progress over the mid-ocean ridge and into the western South Atlantic Ocean. There, three to four years later, they will impinge upon the Brazil Current. It is encouraging to note the close agree-

Cell polarity

From embryo to axon

Melissa M. Rolls and Chris Q. Doe

Many cell types in our body, ranging from neurons to the epithelial cells that line the lungs and skin, must be polarized to function properly. The same mechanism may establish the polarity of many of these cells.

How are we to make sense of the complexity of our brains? Filled with billions of nerve cells, each making hundreds or thousands of precise connections to other neurons, this organ develops anew in each of us. Not only that, but if you were to take a closer look at the individual cells, you would see that they come in thousands of stunning shapes. But there is a theme that emerges from studying these shapes. In vertebrates, most neurons have a single long protrusion, the axon, that is specialized to transmit signals to other cells, as well as many shorter, branched protrusions called dendrites that are specialized to receive signals (Fig. 1). Perhaps understanding this fundamental polarity would be a good starting point for understanding the brain's complexity. Over the past few years, progress has been made in defining how axons and dendrites differ. But less is known about how they are initially established.

ment of the RAFOS float speeds to results of the MODAS — Modular Ocean Data Assimilation System — model, which assimilates satellite altimetric measurement of sea-level variability.

There is still a great deal to learn about the Agulhas valve, and its variation under different climatic conditions. Ensuring that it is properly represented in global ocean and climate models remains a daunting challenge. But this collection of papers² shows how the brotherhood of observers armed with new tools, aided by satellite-based remote sensing, and modellers with their increasingly realistic simulations, can take us forward.

Arnold L. Gordon is at the Lamont–Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA.

e-mail: agordon@ldeo.columbia.edu

1. Bang, N. D. *S. African Geog. J.* **52**, 67–76 (1970).
2. Richardson, P. L. (ed.) *Deep-Sea Res. II* **50**, 1–319 (2003).
3. Matano, R. P. *J. Phys. Oceanogr.* **26**, 2267–2279 (1996).
4. Reason, C. J. *J. Clim.* **14**, 2769–2778 (2001).
5. Gordon, A. L. *J. Geophys. Res.* **91**, 5037–5046 (1986).
6. Gordon, A. L. in *Ocean Circulation and Climate* (eds Siedler, G., Church, J. & Gould, J.) 303–314 (Academic, 2001).
7. de Ruijter, W. A. et al. *J. Geophys. Res.* **104**, 20885–20910 (1999).
8. Gordon, A. L., Weiss, R. F., Smethie, W. M. Jr & Warner, M. J. *J. Geophys. Res.* **97**, 7223–7240 (1992).
9. Sloyan, B. M. & Rintoul, S. R. *J. Phys. Oceanogr.* **31**, 1005–1030 (2001).
10. Gordon, A. L. & Fine, R. A. *Nature* **379**, 146–149 (1996).
11. Speich, S. et al. *Geophys. Res. Lett.* **29**, 10–13 (2002).
12. Beal, L. M., Field, A. & Gordon, A. L. *J. Geophys. Res. Oceans* **105**, 8549–8564 (2000).
13. Weijer, W., de Ruijter, W. P. M. & Dijkstra, H. A. *J. Phys. Oceanogr.* **31**, 2385–2402 (2001).

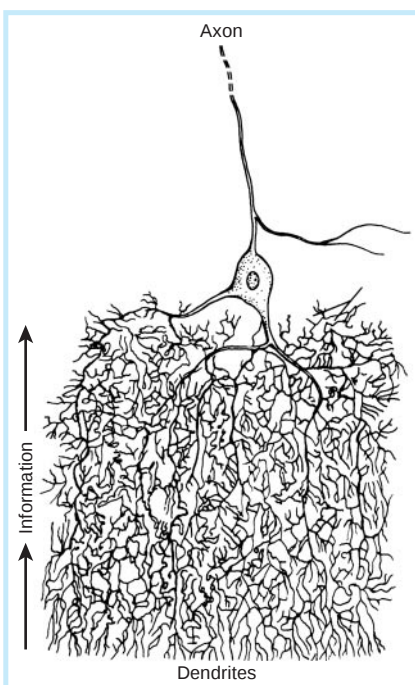


Figure 1 Neurons are highly polarized. This Purkinje neuron, from the cerebellum region of the brain, has a single long axon and a highly branched dendrite. (Modified from ref. 8.)

including tau and CRMP-2 (ref. 3). Dendrite microtubules are organized in part by the motor protein MKLP1 (ref. 4), which is not known to regulate polarity in other cells.

A general conclusion from these data could be that neurons use 'universal' actin-regulating proteins to initiate cell-shape changes, and use microtubule-regulating proteins to generate or maintain differences in axon and dendrite morphology. Is, then, the use of these actin-remodelling proteins the only shared feature in the establishment of polarity in neurons and other cell types?

One place to start looking for common regulators of cell polarity is the Par proteins, first identified in the tiny worm *Caenorhabditis elegans* for their ability to determine the

polarity of the fertilized egg. Par-3 and Par-6 in particular seem to be part of an evolutionarily conserved protein complex that controls polarity in a range of cell types^{5,6} (Fig. 2). In worms these two proteins localize to the anterior of the fertilized egg, and coordinate aspects of cell polarity that ultimately distinguish the head and the tail of the adult wiggling worm. In organisms from mammals to fruitflies, Par-3 and Par-6 are distributed asymmetrically in epithelial cells (which line the stomach, lungs and skin, for instance). Here, the Par proteins are distributed to the cellular apical side — which faces the outside world, in these examples the air or stomach contents — helping to distinguish it from the basal surface. In addition, Par-3 and Par-6 control the polarity of neural precursors in many organisms.

Shi *et al.*¹ have now looked at the role of Par-3 and Par-6 in determining the polarity of neurons from the hippocampal brain region in rats. When such neurons are cultured they initially extend several similar protrusions, called neurites. One neurite then grows rapidly and takes on the characteristics of an axon (notably long length and the presence of tau protein), and later the remaining neurites mature into dendrites⁷. Shi *et al.* show that Par-3 and Par-6 are at first present throughout the neuron, but during the initial phase of axon outgrowth they relocate to the tip of the growing axon. Overexpression of either protein causes the cells to develop multiple projections with a length characteristic of axons. These results suggest the exciting possibility that neuronal polarity makes use of the same Par complex that regulates polarity in epithelia, neural precursors and early embryonic cells.

But what exactly do Par-3 and Par-6 do in neurons? Work on these proteins in other systems provides some clues. Par-6 can bind to enzymes of the Rho family, leading to an attractive model in which Par-6 recruits these enzymes to the end of the axon, where they remodel actin filaments and facilitate growth. In many cell types Par-3 and Par-6

localize together with another enzyme, atypical protein kinase C, and Shi *et al.* also provide pharmacological evidence that this enzyme has a role in axon outgrowth. They suggest that this might be due to an effect on microtubule dynamics.

Another interesting set of questions pertains to the very earliest events of axon outgrowth. What triggers the loss of Par-3 and Par-6 from immature neurites and their enrichment in the developing axon? Here studies of other systems provide few clues, because little is known about how Par-3 and Par-6 become localized in any cell type. It is also unknown how the timing of Par-3/Par-6 localization in cultured neurons, which lack normal environmental cues, relates to the situation *in vivo*.

Can we learn anything from the presence of Par-3 and Par-6 in developing axons about the flip side of neuronal polarity: the specification of dendrites? Perhaps these proteins are required for any type of polarized neurite growth, including that of dendrites. It is possible that during dendrite outgrowth, which occurs after the stage looked at by Shi *et al.*, Par-3 and Par-6 relocate to these structures. If, however, Par-3 and Par-6 uniquely define axons, perhaps there are other proteins that have an analogous function in specifying dendrites. Proteins that are localized opposite the Par-3/Par-6 domain in fertilized eggs, neural precursors or epithelia would be obvious candidates. The Staufen protein, for instance, is localized opposite Par-3 and Par-6 in dividing neural precursors and to dendrites in mammalian neurons. This is an appealing idea. But so far we know of no protein that is found opposite Par-3 and Par-6 in all cell types (Fig. 2).

For now, the key finding is that neurons may exploit a core complex of proteins that is used to generate polarity in many different cell types and many different organisms. Neurons certainly use unique proteins to generate their highly polarized morphology — one that underlies our ability to read News and Views articles and to design elegant experiments — but there is no doubt that many future experiments will build on this newly discovered link between axons, epithelia and embryos.

Melissa M. Rolls and Chris Q. Doe are at the Institute of Neuroscience and the Howard Hughes Medical Institute, University of Oregon, Eugene, Oregon 97403, USA.

e-mail: cdoe@uoregon.uoregon.edu

- Shi, S.-H., Jan, L. Y. & Jan, Y.-N. *Cell* **112**, 63–75 (2003).
- Bradke, F. & Dotti, C. G. *Curr. Opin. Neurobiol.* **10**, 574–581 (2000).
- Fukata, Y., Kimura, T. & Kaibuchi, K. *Neurosci. Res.* **43**, 305–315 (2002).
- Baas, P. W. *Neuron* **22**, 23–31 (1999).
- Doe, C. Q. & Bowerman, B. *Curr. Opin. Cell Biol.* **13**, 68–75 (2001).
- Ohno, S. *Curr. Opin. Cell Biol.* **13**, 641–648 (2001).
- Craig, A. M. & Bankier, G. *Annu. Rev. Neurosci.* **17**, 267–310 (1994).
- Kandel, E. R., Schwartz, J. H. & Jessell, T. M. (eds) *Principles of Neural Science* 3rd edn (Elsevier, New York, 1991).

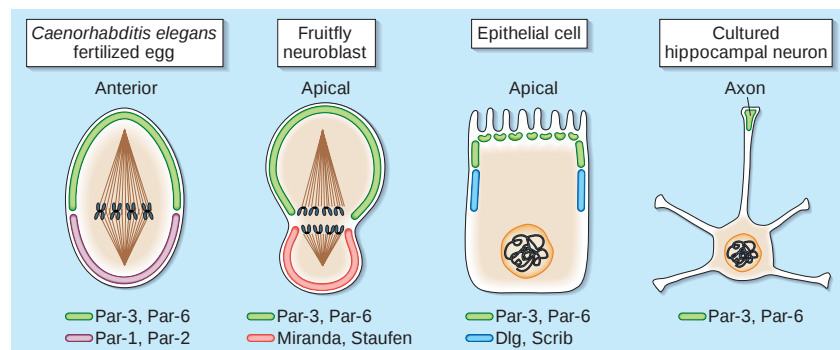


Figure 2 Par proteins produce polarity. The Par-3 and Par-6 proteins establish polarity in various cell types, including the fertilized eggs of the worm *Caenorhabditis elegans*, fruitfly neural precursors (neuroblasts), epithelial cells from fruitflies and mammals, and — as shown by Shi *et al.*¹ — cultured hippocampal neurons. Par-3 and Par-6 show a polarized localization in these cell types and restrict other (cell-type-specific) proteins to different regions of the cell.

Cell biology

Prion protein, too, shall pass

J. Cell Biol. doi:10.1083/jcb1604iti2 (2003)

Part of the job of a cell's endoplasmic reticulum is to act as a transport system for proteins, which are recognized by their so-called signal sequence and enter the system through a dynamic portal, the translocon. Many proteins are processed perfectly well through a translocon that has just a few, well-recognized components. But what about those, such as prion proteins, that are not?

Ryen D. Fons and colleagues have looked into the case. They find that the prion protein has an especially unusual signal sequence, and that effective initiation of prion translocation requires the translocon to have an extra component, the 'translocon-associated protein' (TRAP) complex. It turns out that TRAP also stimulates the translocation of other proteins in a signal-sequence-dependent manner.

One explanation proposed by Fons *et al.* is that TRAP is required on the inside of the endoplasmic reticulum to stabilize the signal sequence and the translocon when they have failed to form their usual intimate association. Given the involvement of prion protein in neurodegenerative disorders, they also point out that the special requirements for translocation could influence the course of prion disease.

Mirella Buccì

Planetary science

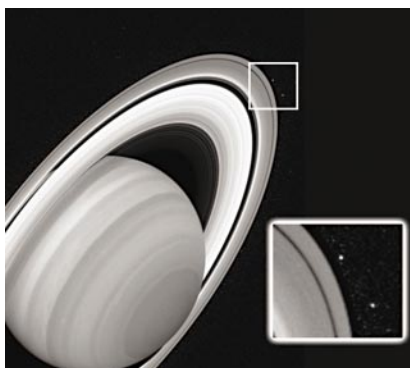
Moons on the wander

Icarus doi:10.1016/S0019-1035(02)00050-7 (2003)

Prometheus and Pandora, two tiny moons of Saturn, have been straying: chaotic interaction between them could be responsible for the perturbation in their orbits. The moons flank the F ring, which lies beyond Saturn's main ring system and is itself dynamically complex.

The two moons are among the smaller of Saturn's 18 satellites (see picture). They were discovered by the Voyager space probe in 1980 and Voyager observations were used to calculate their orbits. Between 1994 and 2000, Richard G. French and colleagues detected gradual changes in the pair's orbits using the Hubble Space Telescope. The variation was in keeping with those of Saturn's other moons, but from images taken in 2001 and 2002 French *et al.* found that Prometheus's orbit had suddenly become 330 metres wider, whereas Pandora's had shrunk by 420 metres.

The authors conclude that these changes have a common origin. They believe that the moons are probably exchanging orbital energy as they hurtle around Saturn, and



Dynamic dancers: Saturn and (inset) its quickstepping satellites Prometheus and Pandora at the outer edge of the planet's rings.

that a state of resonance results in chaotic behaviour. Such behaviour is theoretically possible, but has not previously been observed in the Solar System over such short timescales.

Tom Clarke

Soft matter

Paste inside

Phys. Rev. Lett. 90, 068303 (2003)

Between solid and liquid there exists a host of alluring materials: toothpaste, oil paint, wet cement, peanut butter. Yet these ubiquitous substances are poorly understood. Michel Cloitre and colleagues have now forged a link between the large-scale flow behaviour — the 'spreadability' — of pastes and the motions of their constituent particles.

Cloitre *et al.* made pastes from crosslinked polymer gels and added tiny polystyrene spheres that scatter light, to trace the microscopic structure and motions in the material. In a liquid, the particles gradually diffuse through space. In a glass, this diffusive motion has been slowed to a standstill: the particles become trapped in some random configuration. It seems that pastes are rather like glasses in which the particles are stuck in rubbery traps — elastic cages formed from their neighbouring particles, in which they can rattle around but not escape.

If the paste is stressed, the particles can spring their cages and find a new arrangement. Cloitre *et al.* find that these rearrangements have a characteristic duration, and that this 'microscopic' timescale controls the bulk flow behaviour.

Philip Ball

Immunology

Vaccinia on the Toll road

J. Exp. Med. 197, 343–351 (2003)

Toll-like receptors are proteins that recognize bacterial and viral products, and are essential for kick-starting our immune

response to microbial invaders. Mary T. Harte *et al.* now propose that vaccinia virus — used to vaccinate against smallpox — can obstruct the path from receptor activation to immune fortification.

In the event of a microbial infection, the Toll-like receptors — ten of which have been identified to date — can activate the transcription factor NF- κ B, which turns on genes involved in immune defence. The main track from Toll-like receptor to NF- κ B involves a lengthy sequence of adaptor proteins and enzymes. En route one finds the proteins MyD88 and TOLLIP (Toll-interacting protein), both of which attach to Toll-like receptors. This complex recruits the enzymes IRAK or IRAK2, which phosphorylate themselves and then bind the adaptor protein TRAF6. By way of at least two more stepping-stones, TRAF6 finally activates NF- κ B.

Harte *et al.* find that the vaccinia virus protein A52R may disrupt this pathway by binding the enzyme IRAK2 and the adaptor TRAF6. Moreover, mice infected with a virus that lacks the A52R gene seem less ill than mice suffering from a normal vaccinia virus infection, indicating that A52R may indeed contribute to viral virulence.

Marie-Thérèse Heemels

Imaging

Nanoprobes on the blink

Appl. Phys. Lett. 82, 1102–1104 (2003)

Figuring that a light bulb is easier to see in daylight if it is flashing, Jeffrey N. Anker and Raoul Kopelman have made nanometre-scale fluorescent probes that blink on and off. This modulation helps to separate the probe signal from the background fluorescence.

Particles ranging from many nanometres to a few micrometres in size are becoming widely used as tags and probes in biology and chemistry: for example, to label particular molecules, cells or tissue types, for detecting pathogens and for tracking combinatorial libraries of molecules. Typically these nanoprobes signal their location and identity by emitting fluorescent light.

But other objects in the sample or instrument may also fluoresce, masking the signal. To create a distinctive optical fingerprint in dye-labelled polystyrene microspheres, Anker and Kopelman use spheres embedded with ferromagnetic material. They coat one hemisphere with gold or aluminium, making it dark. When a magnetic field makes the spheres rotate, they blink on and off as the 'dark side' revolves in and out of view. This modulation makes the signal detectable even against a fluorescent background that is ten times brighter.

Philip Ball

A pre-industrial source of dioxins and furans

Domestic burning of coastal peat has produced these nasty pollutants for millennia.

Dioxins and furans (together referred to as dioxins) are ubiquitous, toxic and environmentally persistent organochlorine compounds, which have until now been assumed to be by-products of the organochlorine-based industries that underwent rapid expansion during the 1950s¹. Here we show, however, that the burning of coastal peat was a significant source of dioxins long before the industrial revolution. As coastal peats were burned

extensively over several millennia in the British Isles and beyond, our discovery shows that it is not only modern society that is responsible for this widespread pollution.

Dioxins are contaminants that are produced in the manufacture of polychlorinated biphenyls (PCBs) and organochlorine pesticides, and also by inefficient burning of organic carbon in the presence of chlorine². They have been found in archived soil from the nineteenth century, but the source of

these dioxins has been a mystery³. Wood burning (for fuel and in forest fires) has been implicated as a pre-industrial source of these compounds, but it is not significant because wood is low in chlorine⁴.

As trees generally do not grow well in northern-Atlantic coastal locations, the main fuel resource there used to be peat, which has a high chloride content as a result of maritime-dominated deposition. We therefore tested the smoke and ash from burning peat for contamination with dioxins. The total dioxin concentration (sum of the congeners containing four or more chlorine atoms) in unburned peat (containing 149 mg kg⁻¹ of water-extractable chloride) was 114 ± 14 ng kg⁻¹ (n=3); upon burning, this produced a total of 643 ± 61 ng kg⁻¹ (n=3) in the smoke and the burned peat (expressed per unit mass of peat burned). The amount of dioxins generated by peat burning is therefore significantly greater (*P* < 0.001) than was originally present in the unburned peat.

During the eighteenth and nineteenth centuries, the human population of the highlands and islands of Scotland was 260,000–420,000, most of whom were west-coast crofters (subsistence farmers)⁵. The most detailed anthropological reports for the Western Isles of Scotland are for the Isle of Hirta, St Kilda⁶; we therefore used this site in our study. In the eighteenth and nineteenth centuries, the average household on Hirta contained four to five people⁶, and peat consumption per household can be estimated to be more than 60 m³ per annum (the volume required for over 20 peat-storage huts per household⁶, each of internal volume about 3 m³). Peat has a dry-bulk density of 0.2–0.4 tonnes m⁻³ (ref. 7), which increases upon drying when stacked for fuel consumption.

If we assume that there were four members of each household, and that each household burned 20 tonnes of coastal peat per year, then the highlands and islands could have produced about 1 kg yr⁻¹ of dioxin in total. This is a remarkable figure for this remote region of Britain. Modern estimates of dioxin production by domestic coal combustion are 5.1 kg yr⁻¹ for the entire United Kingdom, with municipal waste incinerators, the main dioxin source, producing 10.9 kg yr⁻¹ (ref. 8).

In many coastal communities, peat ash was collected and added to arable soil to improve its fertility^{6,9}. Hirta, which was evacuated in 1930, has high-chloride peat, and its soil has been deepened over the millennia by the addition of peat ash. We obtained peat ash, preserved in its

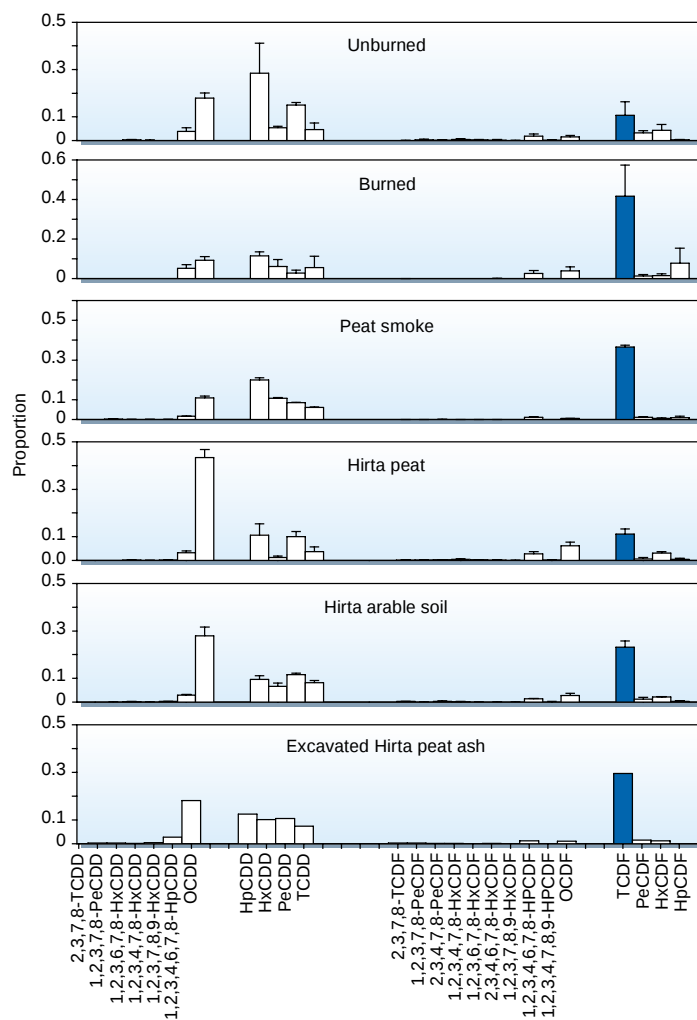


Figure 1 Proportion of dioxin and furan congeners in high-chloride peat (as well as in burned peat and peat smoke) from Gairloch, UK, and in peat, arable soil and peat ash from a black house floor on Hirta. Total dioxins for each matrix (tetra-octa-chloro-dioxins and furans) are (from top to bottom, in ng per kg ± s.e.m.; average of three measurements made at individual sites): 114 ± 14; 133 ± 62; 561 ± 107; 378 ± 146; 217 ± 9; and 60.6 (for Hirta peat ash excavated from a home built between 1830 and 1850, only one sample was available). The total dioxin for peat smoke is expressed on the basis of the quantity of peat burned; all other dioxin totals are expressed as the quantity present per unit dry mass of soil. Blue bars, tetrachlorodibenzofurans. Peat was burned in an oxygenated glass vessel and the smoke trapped on a polyurethane-foam plug; dioxins in all materials were extracted by flushing with hot solvents (Soxhlet). After cleaning the extracts, dioxins and furans were assigned by high-resolution gas chromatography with mass spectroscopy, using ¹³C-labelled dioxins and furans as internal standards. Samples were analysed in an accredited laboratory and met stringent quality-assurance/quality-control requirements.

stratigraphic context and dating from 1830–1850, from excavation of a dwelling⁹. The sum of tetrachlorodibenzofurans (TCDFs) dominates the congener profile of archaeological peat ash, accounting for 30% of all of the dioxins and furans analysed, as in the peat-burning experiments described here (Fig. 1) and in high-chloride lignite-combustion products¹⁰. TCDF levels differed significantly ($P=0.005$, one-way analysis of variance) between the tested matrices in Fig. 1 (the dominant TCDF in the congener profile is a marker for peat/lignite burning).

The environmental persistence of individual congeners varies in different soils¹. Dioxins undergo long-distance global transport¹, and modern deposition explains the presence of dioxins in unburned peat and the high proportion of tetrachlorodibenzo-*P*-dioxin and octachlorodibenzo-*P*-dioxin congeners in these soils compared with the archaeological ash sample (Fig. 1). Even with the superimposition of modern dioxin deposition and the differential loss of

individual dioxins over time, the arable soil of Hirta still carries the signature of past peat ash, with high levels of TCDF compared with fresh peat (Fig. 1). Today, 70 years after the island was evacuated, the dioxin signature accrued over centuries of peat burning is still evident, a reminder that modern humans were not the first to generate large quantities of dioxins.

Andrew A. Meharg, Kenneth Killham

School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK
e-mail: a.meharg@abdn.ac.uk

1. Alcock, R. E. & Jones, K. C. *Environ. Sci. Technol.* **30**, 3133–3143 (1996).
2. Meharg, A. A. & Osborn, D. *Nature* **375**, 353–356 (1995).
3. Alcock, R. E. *et al. Environ. Sci. Technol.* **32**, 1580–1587 (1998).
4. Martinez, M. *et al. Chemosphere* **41**, 1927–1935 (2000).
5. Lynch, M. (ed.) *The Oxford Companion to Scottish History* (Oxford Univ. Press, Oxford, 2002).
6. Harman, M. *An Isle called Hirta* (Maclean, Skye, 1997).
7. *Memoirs of the Soil Survey of Great Britain, Scotland* (HMSO, Edinburgh, 1963).
8. Harrad, S. J. & Jones, K. C. *Sci. Tot. Environ.* **126**, 89–107 (1992).
9. Emery, N. *Excavations on Hirta* (HMSO, Edinburgh, 1996).
10. Thiss, U. *et al. Chemosphere* **31**, 2591–2604 (1995).

Competing financial interests: declared none.

Sociobiology

Worker nepotism among polygynous ants

Insect societies are a prime example of extreme cooperation, but their social life also entails the pursuit of selfish interests by society members¹. Here we show that workers of the ant *Formica fusca* favour their own close kin when rearing eggs and larvae in colonies that are derived from several queens. This nepotistic behaviour indicates that ant workers are able not only to detect kin relationships, but also to pursue their selfish genetic interests if the costs to their colony are not prohibitive.

According to kin-selection theory, social insect workers should selfishly indulge their own evolutionary interests^{1–3}. But evidence

for nepotistic kin discrimination has so far been weak or negative, and may even be precluded by informational constraints and prohibitive costs⁴. We therefore investigated the ability of *F. fusca* workers to discriminate between close and more-distant relatives in colonies that are headed by multiple queens (polygyny). As the workers of this species feed the queens and raise the brood, they may be in a position to manipulate brood composition if they can interpret genetic identity.

The kinship among queens in this species is 0.40 ± 0.14 , and that of workers to the queens is 0.17 ± 0.12 (ref. 5). If workers are more related to one of the colony queens, theory predicts that they should favour her offspring over those of the others^{1,6}; however, if they are equally related to all of the queens, they should

have no incentive to favour the offspring of any particular one^{1,6}.

We studied ten two-queen laboratory colonies⁵ of *F. fusca* to compare within a single cohort the reproductive apportionment at the egg and pupal stages. To determine reproductive apportionment and estimate the relatedness between workers and queens⁷, we genotyped at six micro-satellite loci⁵ an average of 43 eggs, 85 pupae, both queens, and eight workers per colony. Although under laboratory conditions only the worker brood is reared, a queen's genetic representation in this brood accurately reflects her reproductive output (our unpublished results).

We found no positive association between the relatedness of workers and queens and the absolute reproductive apportionment among eggs (Fig. 1a). However, the reproductive share of the queen who was more closely related to the workers increased during brood rearing, and this change was directly proportional to the size of the relatedness difference between the workers and each of the two queens (Fig. 1b). This indicates that workers can discriminate their own kin and selectively favour a brood of closer kin. Although a difference in the viability of eggs⁸ or the presence of diploid males⁹ may cause a pattern of selective disappearance of eggs, neither explains the observed association between worker–queen relatedness and changes in reproductive apportionment.

We conclude that ant workers can apparently discriminate kin accurately and that they capitalize on this ability, thereby enhancing their genetic contribution to future generations¹, even in the presence of several queens. Evidence of nepotism in insects apart from the honeybee (*Apis mellifera*)¹⁰ has not previously been found⁴. This selection for the selfish pursuit of fitness is likely to be a universal feature of societies, but could be modulated by constraints imposed by a lack of accurate information or a reduction in productivity⁴. Power affiliations may be complex in societies in which different parties pursue their individual genetic interests.

**Minttumaaria Hannonen,
Liselotte Sundström**

Department of Ecology and Systematics, University of Helsinki, PO Box 65, 00014 Helsinki, Finland
e-mail: liselotte.sundstrom@helsinki.fi

1. Hamilton, W. D. J. *Theor. Biol.* **7**, 1–16 (1964).
2. Trivers, R. L. & Hare, H. *Science* **191**, 249–263 (1976).
3. Sundström, L. & Boomsma, J. J. *Heredity* **86**, 515–521 (2001).
4. Keller, L. & Reeve, H. K. in *Levels of Selection in Evolution* (ed. Keller, L.) 153–176 (Princeton Univ. Press, New Jersey, 1999).
5. Hannonen, M. & Sundström, L. *Ethology* **108**, 961–973 (2002).
6. Reeve, H. K. & Keller, L. *Annu. Rev. Entomol.* **46**, 347–385 (2001).
7. Queller, D. C. & Goodnight, K. F. *Evolution* **43**, 258–275 (1989).
8. Keller, L. & Ross, K. G. *Science* **260**, 1107–1110 (1993).
9. Cook, J. & Crozier, R. H. *Trends Ecol. Evol.* **10**, 281–286 (1995).
10. Page, R. E. Jr, Robinson, G. E. & Fondrk, M. K. *Nature* **338**, 576–579 (1989).

Competing financial interests: declared none.

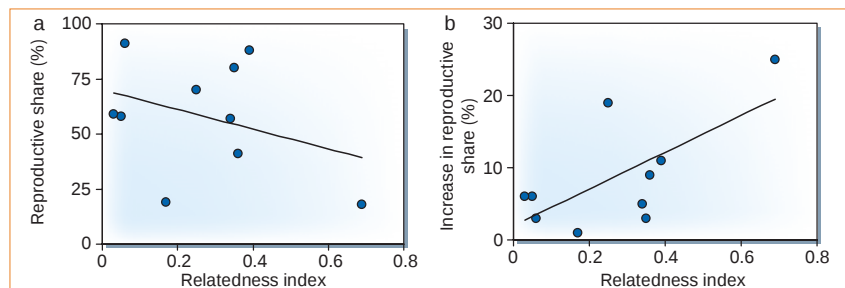


Figure 1 Nepotism in the ant *Formica fusca*. **a**, Absolute reproductive share during brood development among a queen's eggs as a function of her relatedness to the colony workers (correlation coefficient, $r_s = -0.35$, $n = 10$, $P = 0.32$). **b**, Increase in a queen's reproductive share as a function of her relatedness to workers. Correlation between the two variables was highly significant ($r_s = 0.67$, $n = 10$, $P = 0.01$; one-sided test). To exclude any other factor relating to queens' fecundity⁸, the change in reproductive apportionment was quantified by first regressing the reproductive apportionment of each queen among the matured brood on her apportionment among the eggs, and then using the positive residuals. The relative kin value of a queen (relatedness index) is calculated as $r_{wq+} - r_{wq-}$, where r_{wq+} is the relatedness of workers to the queen that increased her reproductive share, and r_{wq-} is the relatedness of workers to the queen whose share decreased.

Audiovisual perception

Implicit estimation of sound-arrival time

In perceiving the sound produced by the movement of a visible object, the brain coordinates the auditory and visual input^{1–3} so that no delay is noticed even though the sound arrives later (for distant source objects, such as aircraft or firework displays, this is less effective). Here we show that coordination occurs because the brain uses information about distance that is supplied by the visual system to calibrate simultaneity. Our findings indicate that auditory and visual inputs are coordinated not because the brain has a wide temporal window for auditory integration, as was previously thought, but because the brain actively changes the temporal location of the window depending on the distance of the visible sound source.

Seven subjects with normal vision and hearing were presented through headphones with a burst of white noise (90 decibels

sound-pressure level, 10-ms duration, with 4-ms rise and fall times), the spectrum of which had been processed (by using head-related transfer functions) to simulate an external sound from a frontal direction. Brief light flashes (10 ms) were produced by an array of five green light-emitting diodes (LEDs) at different distances from the subjects (1–50 m; Fig. 1). The intensity of the light flash was 14.5 candelas per square metre at a viewing distance of 1 m, and was increased in proportion to the square of the viewing distance for the other distances in order to produce consistent intensity at the eye. The difference in onset times between the sound and light stimuli was varied randomly from –125 ms to 175 ms in steps of 25 ms.

Subjects were instructed to look at the centre of the LED array and to imagine that the LEDs were the source of both light and sound, while listening to the sound directly from the sound source. To eliminate possible bias effects, we used a two-alternative forced-choice task to measure subjective simultaneity: in this task, observers judged whether the light was presented before or after the sound. Twenty responses were obtained for each condition. To determine the stimulus-onset asynchrony that corresponded to subjective simultaneity, we estimated the 50% point (the point of subjective equality) by fitting a cumulative normal-distribution function to each individual's data using a maximum-likelihood curve-fitting technique.

When the LED array was 1 m away, the point of subjective equality occurred at a sound delay of about 5 ms; however, the sound delay at this point increased with viewing distance ($P < 0.001$; Fig. 1a, b). This increased delay was roughly consistent with the velocity of sound (about 1 m per 3 ms at sea level and room temperature), so the point of subjective equality increased by about 3 ms with each 1-m increase in distance. This relationship was consistent at least up to a distance of 10 m.

Our results show that the brain probably takes sound velocity into account when judging simultaneity. However, it takes about 120 ms for sound to travel 40 m, and we found that the threshold for detecting the sound delay was 106 ms at a viewing distance of 40 m, so active compensation is likely to operate only for shorter distances than this.

We have shown that the brain takes sound velocity into account when integrating audiovisual information. The brain can therefore integrate audiovisual information over a wide range of temporal gaps, and correctly match sound and visual sources.

Yoichi Sugita*, Yôiti Suzuki†

*National Institute of Advanced Industrial Science and Technology, Neuroscience Research Institute, Teragu-hakusan 1497-1, Tsukuba 300-4201, Japan
e-mail: y.sugita@aist.go.jp

†Research Institute of Electrical Communication

and Graduate School of Information Sciences, Tohoku University, Katahira, Aoba-ku, Sendai 980-8577, Japan

1. Sekuler, R., Sekuler, A. B. & Lau, R. *Nature* **385**, 308 (1997).
2. McDonald, J. J., Teder-Sälejärvi, W. A. & Hillyard, S. A. *Nature* **407**, 906–908 (2000).
3. Shams, L., Kamitani, Y. & Shimojo, S. *Nature* **408**, 788 (2000).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Sociology

The puzzle of human cooperation

Humans often defy rational-choice theory by cooperating in simple dilemma games¹, a paradox that has been explained by theories of kin selection², reciprocal altruism³ and indirect reciprocity (reputation)⁴. Fehr and Gächter claim that human cooperation remains an evolutionary puzzle because people will cooperate with genetically unrelated strangers, often in large groups, with people whom they may not meet again, and without any gain in reputation ('strong reciprocity')⁵ — that is, when existing theories do not seem to apply. However, we argue that those theories are rejected for the wrong reasons and that the paradox may therefore be imaginary. This has implications for whether punishment is crucial to promoting cooperation^{5–8}.

First, strong reciprocity is not necessarily a puzzle because altruistic tendencies need not reflect contemporary selective pressures. Rather, they may reflect motivations that evolved during the past 5–7 million years, in situations that were very different from those presupposed in the puzzle (small groups with relatedness greater than random, individuals well known to each other, interactions likely to be repeated, and people organized in hierarchies). The puzzle disappears if human brains apply ancient tendencies to cooperate that persist in newer environments, even if they are maladaptive (heuristic rules that violate expected utility often make sense for common tasks in our evolutionary history).

Accordingly, kin selection, reciprocal altruism and indirect reciprocity need not explain why altruism fails to conform to rationality theory today; rather, they explain why it became ingrained in our brains in the past. Today, humans distinguish and favour kin, or select partners with whom to repeat games, but this does not conflict with those explanations (it supports their legacy in our evolution). The cost of errors may be high but they are recent, so it is unclear whether selection has been strong (or fast) enough to realign humans' strategic behaviour specifically to deal with modern problems — evidence for systematic deviation from rational solutions indicates that it has not. Such deviations occur across cultures¹, suggesting an ancient and/or common

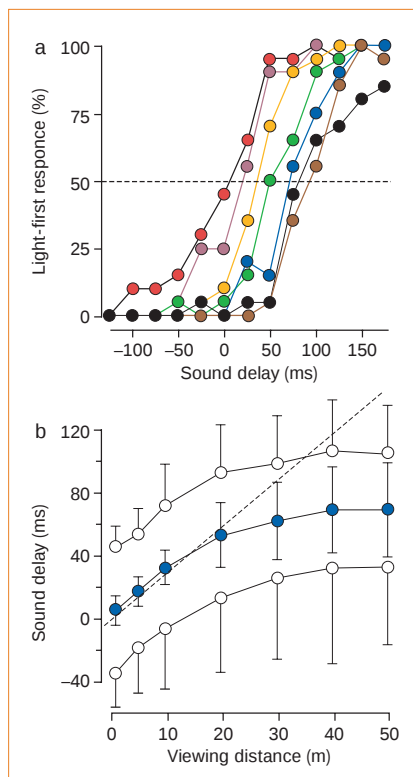


Figure 1 Synchrony in audiovisual perception. **a**, Representative results from one observer. The percentage of light-first response for each viewing distance is plotted against sound delay (stimulus-onset asynchrony). Different colours represent results for different viewing distances (red, pink, yellow, green, blue, brown and black correspond to 1, 5, 10, 20, 30, 40 and 50 m, respectively). Dashed line indicates the 50% point, which corresponds to subjective simultaneity. **b**, Points of subjective equality (filled circles) plotted against viewing distance. Hollow circles, plots of 25% (bottom curve) and 75% (top) of light-first response indicate the threshold for detecting asynchrony. Dashed line represents the real sound-arrival time.

origin, rather than a fine-tuning to varied modern circumstances.

Second, the punishers in laboratory experiments such as Fehr and Gächter's are anonymous⁵, so potential extra costs resulting from retaliation (of any sort) by victims are ruled out. Anonymity is unrealistic among early human groups: vigilantes would have to confront defectors to punish them, which incurs risk, and punishment among group members gives rise to grudges and reprisals, which undermine future cooperation. Although groups may be willing to punish individual defectors, people in one-on-one situations may not accept the personal cost of punishment (for example, they are often unwilling to intervene in criminal acts or to testify in trials for fear of retaliation).

Third, the problem remains of what prevents the occurrence of second-order free-riders, who cooperate for the public good but defect from bearing the cost of punishment⁹. Fehr and Gächter's results suggest that this is not a problem, as a core of people willingly incur personal costs to administer punishment, motivated by anger (although it is unclear whether they would act on it if they were not anonymous).

Alternative solutions are that punishment may come from an external institution, or it is not costly, or is administered to both defectors and individuals who fail to punish defectors. These alternatives have been criticized⁹ but, if punishment is important, we suggest that an important source has been overlooked. The 'external' solution has been rejected because cooperation is prevalent in pre-industrial human groups, despite the absence of the enforcing institutions of modern states⁹ — but this ignores religion, a feature of all human societies. Religions share taboos and codes of conduct that often promote cooperation for the public good and threaten supernatural punishment for those who do not follow these codes. Followers fear the personal consequences of defecting, and may be prepared to be altruistic if they believe that those who are not will be punished (now or in an afterlife).

A belief in supernatural punishment may arise as an abstract product of culture but thereafter become subject to selection (that is, the argument does not rely on invoking evolutionary origins for religious beliefs). But such beliefs could have evolved, either by group selection⁸ or natural selection — as with the 'green beard' effect, individuals who signal common attributes can cooperate selectively with each other and thus outperform others. Groups with costly religious beliefs that signal commitment and loyalty outlive non-religious groups as a result of improved cooperation¹⁰. A 'stick' may be a good way to coerce people into cooperating, but the 'carrots' of kin cooperation, reciprocal altruism, reputation-

building and religion have been crucial alternatives over our long evolutionary history, with a legacy that pervades today.

Dominic D. P. Johnson*, Pavel Stopka†, Stephen Knights‡

*Olin Institute for Strategic Studies, Harvard University, Cambridge, Massachusetts 02138, USA
e-mail: dominic@post.harvard.edu

†Department of Zoology, Charles University, 128 44 Prague 2, Czech Republic

‡Department of Economics, University of Oxford, Oxford OX1 3UQ, UK

1. Henrich, J. *et al.* *Am. Econ. Rev.* **91**, 73–78 (2001).
2. Hamilton, W. D. *J. Theor. Biol.* **7**, 1–52 (1964).
3. Trivers, R. L. *Q. Rev. Biol.* **46**, 35–57 (1971).
4. Alexander, R. D. *The Biology of Moral Systems* (Hawthorne, Aldine, New York, 1987).
5. Fehr, E. & Gächter, S. *Nature* **415**, 137–140 (2002).
6. Clutton-Brock, T. H. & Parker, G. A. *Nature* **373**, 209–216 (1995).
7. Ostrom, E., Walker, J. & Gardner, R. *Am. Polit. Sci. Rev.* **86**, 404–417 (1992).
8. Sober, E. & Wilson, D. S. *Unto Others: The Evolution and Psychology of Unselfish Behaviour* (Harvard Univ. Press, Cambridge, Massachusetts, 1998).
9. Henrich, J. & Boyd, R. *J. Theor. Biol.* **208**, 79–89 (2001).
10. Sosis, R. *Cross-Cult. Res.* **34**, 71–88 (2000).

Fehr and Gächter reply — The claim by Johnson *et al.* that human cooperation in social-dilemma games violates rational-choice theory is not justified¹. If people have altruistic aims, altruistic behaviour is a rational means by which to achieve their proximate goals. From an evolutionary viewpoint, we need to explain why humans are often altruistic by strong reciprocity^{2–4}. Although kin selection, reciprocal altruism and indirect reciprocity explain relevant forms of human cooperation^{5–7}, they do not ultimately explain strong reciprocity⁸.

Kin selection would account for strong reciprocity if human behaviour were driven by rules that do not distinguish between kin and non-kin. But humans, like other primates, distinguish cognitively and behaviourally between the two^{5,9}, and generally feel stronger emotions towards kin. Likewise, reciprocal altruism could account for strong reciprocity if humans' behavioural rules did not depend on the probability of future interactions with potential opponents. But humans can distinguish long-term partners from people with whom future interaction will be less likely ('strangers'), and will cooperate more if they anticipate that interaction will be frequent⁶. Emotional responses may also be stronger towards a long-term partner than towards a 'stranger' (our unpublished results).

Reputation-based ultimate theories could account for strong reciprocity if our behavioural rules did not depend on our actions being observed by others. However, if reputation formation is ruled out, cooperation breaks down, whereas it flourishes if subjects gain in reputation⁷.

Early humans whose behaviour was fine-tuned to respond to kin or non-kin, partners

or strangers, and gaining in reputation, probably had an evolutionary advantage because, contrary to common belief, they faced interactions where the probability of future encounters was sufficiently low as to make defection worthwhile. Ethnographic evidence indicates that humans had many encounters with individuals with whom they had little future interaction⁸. In addition, the costs of mistakenly treating unrelated individuals as kin, or treating strangers as partners, were high — for instance, a lack of vigilance with strangers could be fatal. Because of these costs, individuals who could adjust their behaviour to suit the their opponent's characteristics had greater fitness. The problem with any theory claiming that strong reciprocity is maladaptive in modern circumstances is that individuals understand the risks of exploitation in interactions with non-kin and strangers, and behave accordingly. An evolutionary explanation of strong reciprocity is needed that does not assume that individuals are maladapted^{2,3}.

A proximate mechanism of belief in supernatural punishment does not solve the evolutionary puzzle. How could such beliefs evolve if those who did not hold them defected and hence gained an advantage? Laboratory experiments do not support the claim that religion is important for cooperation. If other people in the group are expected to defect, then almost everyone else — religious or not — will defect too¹⁰. Moreover, in almost all religions, non-believers have been ostracized and have faced worldly punishment.

We do not agree that anonymity is a problem in the experiment: it rules out other, less costly forms of social punishment that are available in non-anonymous situations, such as workers' hostility towards strike-breakers and people's hostility towards wartime deserters. If non-anonymous punishment were lessened by being more costly, this could be just another example of how remarkable humans are at fine-tuning their behaviour to suit their circumstances.

Ernst Fehr*, Simon Gächter†

*Institute for Empirical Research in Economics, University of Zürich, 8006 Zürich, Switzerland
e-mail: efehr@iew.unizh.ch

†University of St Gallen, FEW-HSG, 9000 St Gallen, Switzerland

1. Gintis, H. *Game Theory Evolving* (Princeton Univ. Press, 2000).
2. Henrich, J. & Boyd, R. *J. Theor. Biol.* **208**, 79–89 (2001).
3. Gintis, H. *J. Theor. Biol.* **206**, 169–179 (2000).
4. Fehr, E. & Gächter, S. *Nature* **415**, 137–140 (2002).
5. Silk, J. B. *Am. Anthropol.* **82**, 799–820 (1980).
6. Gächter, S. & Falk, A. *Scand. J. Econ.* **104**, 1–25 (2002).
7. Milinski, M., Semmann, D. & Krambeck, H. *J. Nature* **415**, 424–426 (2002).
8. Fehr, E. & Henrich, J. in *The Genetic and Cultural Evolution of Cooperation* (ed. Hammerstein, P.) (MIT Press, Cambridge, Massachusetts, in the press).
9. Tomasello, M. & Call, J. *Primate Cognition* (Oxford Univ. Press, New York, 1997).
10. Fischbacher, U., Gächter, S. & Fehr, E. *Econ. Lett.* **71**, 397–404 (2001).

Fire science for rainforests

Mark A. Cochrane*†

* Center for Global Change and Earth Observations, Michigan State University, 1405 S. Harrison Road, Room 101, East Lansing, Michigan 48823-5243, USA

† Instituto do Homem e Meio Ambiente da Amazônia (IMAZON), Caixa Postal 1015, Belém-PA, CEP 66017-000, Brazil

Forest fires are growing in size and frequency across the tropics. Continually eroding fragmented forest edges, they are unintended ecological disturbances that transcend deforestation to degrade vast regions of standing forest, diminishing ecosystem services and the economic potential of these natural resources. Affecting the health of millions, net forest fire emissions may have released carbon equivalent to 41% of worldwide fossil fuel use in 1997–98. Episodically more severe during El Niño events, pan-tropical forest fires will increase as more damaged, less fire-resistant, forests cover the landscape. Here I discuss the current state of tropical fire science and make recommendations for advancement.

Fires in the tropics generally bring to mind decades of media images of deforestation fires where slash-and-burn practices are used to convert lush rainforests into agricultural lands. However, beyond these intentional fires looms the burning of vast regions of standing rainforests as an unintended consequence of current land-use practices. Quantitative estimates of area burned are sporadic, incomplete or non-existent for most tropical countries but are nevertheless shocking. Within southeast Asia and Latin America, out-of-control fires burned more than 20 million hectares in 1997–98 (refs 1, 2). This is equivalent to half of California and is far from a complete accounting of unintentional burning in the tropics, because burning in many regions has not been well quantified.

Some perspective is necessary to grasp the scope of the problem faced in the tropics. The fire season of 2000 was one of the worst on record in the United States, with nearly 3.4 million hectares burned and over 1.4 billion US dollars in fire suppression costs (see the National Interagency Fire Center website, <http://www.nifc.gov/stats/>). By way of contrast, in 1997–98, Indonesia was faced with 8 million hectares of burning and a fire-fighting budget, including foreign aid, of only \$25 million (ref. 3). The story is similar across the tropics: 3 million hectares burned in Bolivia⁴, 2.5 million hectares burned throughout Central America², and in Brazil, 5 million hectares burned in a single Amazonian state⁵. The ubiquitous nature of fire in the tropics is evident in satellite imagery. Over the years, several satellites have been pressed into service as fire-detection devices (Box 1). These and newer, more adept, satellites show a landscape that appears to burn every year (see <http://earthobservatory.nasa.gov/Observatory/showqt.php3>).

Throughout the last century, great efforts and vast resources have been applied to understanding and managing fire in forests. However, in the tropics, demographic and land-use changes have only recently made fire a matter of serious concern⁶. Consequently, critical knowledge and resources have been lacking. Fire science has come late to the tropics.

With so much existing knowledge about fire science it is fair to ask whether tropical fires merit specific study. Fire itself is simply a mixture of heat, fuel and oxygen. The particular mixture of these three components determines whether and how things burn. Integration of fire with the physical and biological components of the Earth leads to a remarkable amount of complexity in fire behaviour and effects. Add to this the evolutionary, economic and social context under which fire operates and it becomes clear that while the chemistry of fire may be universal, its effects are not. This review attempts to explain the present understanding of fire in tropical rainforests and outline areas needing further development.

Impacts of fire

Ecosystem effects

Burning of tropical evergreen rainforests alters forest composition and structure^{7,8}. Common tree species suffer the greatest total mortality⁹ but rare species are most likely to be locally extirpated⁷. Even 15 years after burning, forests show no evidence of regaining lost species⁹. Prospects for species recovery are diminished because surface fires reduce seed availability by 85% in the litter layer and 60% in the upper 1.5 cm of soil¹⁰, while flowering and fruiting of trees in and near burned forests also decreases¹¹. Conditions heavily weight post-fire regeneration toward wind-borne, light-demanding pioneer species. Unburned forest islands^{7,10} and gallery forests¹² are key seed sources for post-fire recovery; however, recurrent fires rapidly reduce the size and density of surviving unburned forest fragments⁷ and kill regenerating vegetation, further depleting the prospects for recovery of mature forest species¹⁰. Grasses and small vines also invade burned forests, slowing regeneration and increasing forest flammability^{8,13–15}.

Animal responses to fire are poorly understood but are complex and dynamic, with some groups, including small mammals, reptiles and amphibians, increasing in abundance or species richness¹⁶ while other groups decline (for example, birds, insects)^{16,17}. Most animals initially escape the fires but die through starvation or predation if they cannot find shelter and establish new territories².

With millions of hectares of fire-damaged forests spanning the tropics, studies of natural and managed ecosystem recovery are needed. In particular, the rates at which these damaged rainforests recover their near-fire-immune status need to be determined. Potential intervention or management strategies to enhance ecosystem recovery processes should also be assessed.

Emissions

Net annual carbon emissions from tropical wildfires are extremely variable, ranging from 7.5 to 70 Mg ha⁻¹ depending on previous fire and land-use history¹³. Emissions from unplanned tropical peat fires are more severe, possibly exceeding 300 Mg C ha⁻¹ (ref. 18). Few estimates of total atmospheric carbon emissions from tropical forest fires exist but include 4.6 Tg from Mexico¹⁹, 23.2 Tg from Roraima, Brazil²⁰, and 810–2,570 Tg from tropical peat fires in Indonesia¹⁸ for 1997–98 El Niño fires. These three fires alone may have equalled 41% of world fossil fuel emissions²¹. Millions more hectares are known to have burned in southeast Asia¹, Africa, Central and South America². The vast majority of unconsumed biomass from killed trees will also slowly release carbon through decay processes^{1,7}.

Smoke from tropical forest fires causes human suffering and death. Across the tropics, hundreds have lost their lives in smoke-related accidents, including ship, automobile and plane crashes^{1,2}.

Box 1

Fire detection from space

Fire is a chemical reaction that releases energy in the form of heat and light. Burning changes the structure of vegetation and converts fuels into a host of combustion products that can remain in place (such as char and ash) or be transported from the site (such as aerosols). Conditions during and after fire change the spectrum of reflected and emitted energy from an affected area. Earth-observing satellite sensors can potentially make use of these changes to detect fires, quantify areas burned and determine the composition and distribution of smoke plumes.

Ideally, fires should be mapped by low-cost, frequently collected satellite imagery, but determining which tropical forests are burning or have burned has not turned out to be a simple exercise. Fire detection becomes a challenge of balancing missed detections of real fires and erroneous detections of false fires⁸⁰ (such as sun glint, clouds, bare soil surfaces). High-resolution satellites including the Landsat Enhanced Thematic Mapper (ETM+) and Systeme Probatoire d'Observation de la Terre (SPOT) can be used to detect fires but their relatively high cost, and limited spatial and temporal sampling, limit their usefulness for active fire detection⁸⁰. Frequently used satellite systems include the Advanced Very High Resolution Radiometer (AVHRR), the Geostationary Operational Environmental Satellite (GOES) and the Defense Meteorological Satellite Program Operational Line Scanner (DMSP-OLS) satellites because of their frequent coverage of the Earth's surface. Although originally designed for cloud detection, each sensor has different strengths and limitations for fire detection⁸¹. Issues include sensor sensitivity and saturation characteristics, spatial resolution, number of observations per day and processing methods among others². Several algorithms have been used to estimate the area burned with varying results. Threshold and contextual classifiers tend to detect different fires⁸². Within contextual classifications, different methods lead to differing detection capabilities of different types and sizes of fires⁸⁰, which is important in the tropics because deforestation fires are very hot and may smoulder for as much as a week⁶⁹ while forest fires can be of low intensity and low temperature¹³ and fires in open vegetation types, such as cerrado and pastures, can burn intensely but cool quickly⁸³. Where low-intensity fires are widespread over a region, they can even influence discrimination statistics of contextual classifiers in such a way that prevents detection of normally detected intense fires⁸⁰.

Newer satellite sensors are expanding these capabilities in monitoring the extent and frequency of fires⁸⁴. The Tropical Rainfall Measuring Mission—Visible and Infrared Scanner (TRMM—VIRS), may be particularly useful for detecting unusually large or hot fires because of the reduced sensor saturation of its larger sampling area and potential for discriminating fires from other hot surfaces⁸⁵. The Bi-Spectral Infrared Detection (BIRD) mission is currently testing a new generation of infrared array sensors that provide surface temperature, area, and geo-location of fires with a spatial accuracy of 300 m, better analysis capabilities for High Temperature Events (HTE) and the ability to discriminate between smoke and water clouds^{86,87}. The higher saturation levels of Moderate Resolution Imaging Spectroradiometer (MODIS) satellite sensors provide better detection and monitoring capabilities for active fires than previous satellite platforms. In addition, improved spatial and radiometric resolutions of these satellites are useful for investigating aerosols in smoke plumes²⁷ and enhance the accuracy of burned area estimates⁸⁸. Both coarse⁸⁸ and fine⁸⁹ spatial resolution sensors are useful for mapping forest burning in the tropics. Burn signatures are transient in imagery owing to rapid forest regeneration⁹⁰ but may be detectable for up to two years with some techniques⁸⁹. ERS-2 Synthetic Aperture Radar (SAR) imagery is also promising for mapping burn scars in tropical evergreen forests because of its ability to penetrate clouds and smoke⁸⁴; however, detection capabilities can be confounded by moisture, making knowledge of site weather information crucial.

Thousands more died from smoke-related illnesses from forest fires in Indonesia²² and Brazil²³. Health effects depend on the concentration, constituents and length of exposure to smoke²⁴ but include respiratory and cardiovascular difficulties among other illnesses²². The complex mix of particles, liquids and gaseous compounds released depend upon the type and efficiency of burning. Fire emissions have been studied and quantified for deforestation and savanna fires²⁵ but not for tropical forest fires. Emissions may change greatly between day and night because many of the fires then convert from flaming to smouldering combustion¹³. Emissions can be trapped beneath the rainforest canopy, potentially leading to toxic conditions. In addition, smoke plumes from biomass burning release large amounts of aerosols to the atmosphere, a substantial fraction of which is black carbon. In part owing to the black carbon fraction, the mix of aerosols in the tropics reduces the average solar radiation absorbed at the Earth surface by up to 35 W m⁻² and may double the heating of the lower 3 km of the atmosphere²⁶. These changes affect atmospheric stability and cloud formation and can ultimately reduce rainfall over large areas^{27–29}.

Development, deforestation and fire

Development of forested areas has historically been synonymous with deforestation. The advent of slashing and burning tropical forests for agricultural purposes was thousands of years ago³⁰. Periodic clearing and abandonment of tropical rainforests is potentially a sustainable practice, if rotation rates are long enough, but can only support low population densities. Higher population densities, extensive forest clearing and extreme drought conditions of periodic mega-El-Niño events may have resulted in widespread forest fires, frequent and severe enough to have influenced the distribution and migration of indigenous people³¹. In the modern era, tropical forests are the 'frontier' and, as always, fire is the tool of choice for clearing the land. Fire is very useful because it quickly and effectively reduces the biomass of newly cleared forests to nutrient-rich ash that can fertilize crops. Hard-won agricultural lands are managed using fire to keep forest regrowth at bay, reduce pests, and promote forage growth for domestic animals. Fire is needed to create charcoal, reduce trash and debris piles, and for cooking food.

Fire becomes a problem, however, when it escapes from its intended purpose and causes unanticipated economic and environmental damage. Heavy fire use has costs as well as benefits. Escaped fires destroy crops, pasturage, timber, infrastructure (for example, fence posts, buildings and so on), and livestock, at times leading to severe economic hardship for local populations^{2,32}. Ill-timed fires can also spoil newly cleared lands by under- or overburning the slashed trees, resulting in reduced productivity or abandonment. Careless fire use and arson can lead to extensive fires in logged and unlogged forests and reduce the profitability or attractiveness of large-investment, high-profit permanent agriculture.

Fire and tropical forests

The crux of the fire problem in tropical rainforests is not so much the introduction of fire into these ecosystems but the frequency with which they are being burned. Historical records^{1,33,34} and charcoal in soil profiles^{35–38} show that tropical forest fires, even in the wetter forests, are not unprecedented. Fire can be considered endemic but rare in tropical rainforests¹⁴ with return intervals of hundreds if not thousands of years³⁹. Wetter forests burn less frequently but are more vulnerable to fire than drier forests because they have thinner protective layers of bark⁴⁰ and suffer much higher mortality rates from fires. Infrequent fire disturbance has left tropical rainforests evolutionarily ill-adapted to current patterns of burning^{40,41}.

Fire susceptibility in tropical forests occurs largely because of moisture stress, during periods of extensive drought, when normally moist fuels dry out and become potentially flammable. However, closed-canopy evergreen rainforests are remarkably resistant to drought³⁵. Even after months without rain, these forests can

maintain an evergreen canopy and high sub-canopy humidity levels, making sustained combustion impossible^{35,39}. This apparent fire-immunity results from the effective trapping of transpired moisture such that most of the ambient forest humidity is derived from the trees themselves⁴². Resilience to climatic stress through moisture recycling^{43,44}, enhanced by the deep-rooting capacity of many tree species⁴⁵, allowed many tropical forests to persist through severe droughts of previous glacial periods⁴⁶. But despite adaptations to drought conditions, rainforests can and do burn. Both natural and anthropogenic disturbances to forest canopies decrease the ability of forests to maintain moisture, making them more vulnerable to fire. In anthropogenic landscapes, fire is a continual presence, making it only a matter of time until a sufficiently intense drought opens up the forest to fire².

Fire occurrence in tropical forests is largely associated with forest edges^{47–51} and is not randomly located throughout. Although tropical regions have the highest density of lightning strikes⁵² and some associated tree mortality, these events are usually associated with heavy rainfall and rarely lead to forest fires⁵³. Other natural causes of fire such as volcanic eruptions and burning coal seams³⁷ may be regionally important but, as in most locations throughout the world, fire in the tropics is primarily associated with human activity^{34,54} and influence on land cover^{47–51}.

Land cover effects

Tropical land cover is rapidly changing⁵⁵. Road construction in recent decades has provided access for millions of settlers to previously remote and inaccessible forests⁵⁶. Deforestation has necessarily followed. Forest clearance fragments the remaining forests^{55,57}. Resultant forest edges are buffeted by winds and desiccating sunlight. These edge effects lead to structural changes, including increased mortality of trees, decreased living biomass⁵⁷

and increased fuel loads (that is, woody debris)⁵⁸. These changes predispose these forests to fire^{48,49} (Fig. 1).

Selective logging is another major landscape modifier in the tropics. The combination of human access provided by logging roads⁵⁹ and the forest damage caused by logging activities³⁹ make logged forests extremely vulnerable to fire⁶⁰. This vulnerability may last for decades after the logging activities have ceased⁶¹ (Box 2).

Given the juxtaposition of fire-prone forests and fire-dependent agricultural lands, forest fires are almost inevitable. In new frontier areas with relatively low quantities of cleared land, extensive undamaged forests make large forest fires nearly impossible. However, as fire-prone agricultural land development continues, the forest becomes increasingly fragmented and the whole landscape becomes conducive to fire propagation². The forest fires that do occur are often edge-related, moving into forests from deforested lands^{48,51}. These fires can significantly alter fire regimes kilometres from forest edges (Fig. 2). Fire frequency becomes a function of distance from deforested forest edges^{48,62} and fire severity increases with frequency¹³. In the absence of other modifying disturbances, these forests will continue to erode, with isolated fragments collapsing, unless future fires can be prevented⁶³. This is a long-standing process in the African tropics⁴³. Within the Brazilian Amazon alone, 26 million hectares of forest could currently be undergoing this process, leading to the eventual release of 3.9 Pg C (ref. 62).

Fire behaviour

Initial fires in relatively intact rainforests do not seem to be severe. Often progressing as slowly creeping ribbons of flame, 10 cm or so in height, burning little besides fallen leaf litter, these fires are still capable of killing 23–44% of the trees >10 cm in diameter at breast height (DBH)^{7,8,10,15,40,60}. Fire propagation in tropical forests is largely controlled by variations in ambient relative humidity

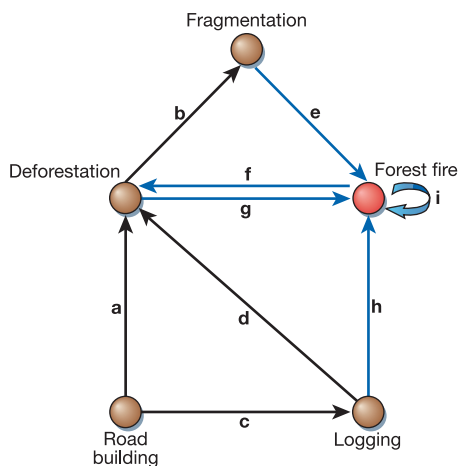


Figure 1 Diagram of interrelationships between tropical land-cover changes and forest fires. Arrows indicate forcing of each node upon others in the system. Blue arrows directly affect forest fire occurrence; black arrows indirectly influence forest fire occurrence. Events **a–i** are: **a**, Road building results in forest access that is strongly associated with deforestation⁵⁶. **b**, Deforestation fragments the remaining forests, increasing amounts of edge^{55,57}. **c**, Road building and paving directly affect transportation costs and area of economic accessibility⁷⁸. **d**, Logging results in limited amounts of deforestation for roads and log landings. Post-logging colonization can increase deforestation⁵⁹. **e**, Forest edges suffer biomass collapse and microclimate changes⁵⁷ making them susceptible to frequent fires⁴⁸. **f**, Repeated forest fires can lead to unintentional deforestation¹³. **g**, Deforestation and pasture/land maintenance fires result in many accidental forest fires⁶². **h**, Logging degrades forests, increasing fire susceptibility^{47,60,61}. **i**, Forest fires can create a positive feedback cycle where recurrent fires become more likely and more severe with each occurrence¹³.

Box 2

Selective logging and fire

Millions of hectares of tropical forests have already been logged and millions more are being logged each year^{61,91,92}. Logging in tropical forests is often selective in that only a few valuable trees are extracted, leaving most of the forest in place. Impacts of selective logging vary with extraction intensity and management practices^{93–95}, but can be substantial. Selectively logged forests can recover to pre-harvest levels of biomass if left undisturbed. However, many forests are revisited several times when loggers return to harvest additional tree species as regional timber markets develop⁹³. These forests become very degraded and may have 40–50% of the canopy cover destroyed during these logging operations⁹³. Selective logging increases fire susceptibility^{39,60}, and can catalyse fire use and deforestation by opening up unoccupied and protected lands to colonization⁵⁹.

Fire severity in logged forests can be high^{2,8,61} owing to the large amount of available fuel in the form of slash piles and collateral damage caused by the logging operations^{39,93}. When weighted by area, logged forests near forest edges may not be more likely to burn than unlogged edge-forests but burn severity and fire spread rates can be higher, leading to deeper penetration and increased damage by fire (refs 34, 50, 96; see also http://www.fire.uni-freiburg.de/se_asia/background/sea_5.html). In landscapes heavily affected by logging and fire, the relationship between fire frequency and distance from deforested edges can decrease or disappear completely because the entire forest will have similar fire-return intervals, often less than 10 years (ref. 48). Of deforestation, fire use or selective logging, none by itself is likely to create severe fire problems in tropical landscapes but the synergy and interaction between these land-uses when present together can promulgate fire throughout the landscape and rapidly degrade forests^{2,47–50}.

(Box 3)^{13,39,64}. Although fire intensity is very low ($4\text{--}55\text{ kW m}^{-1}$), slow spread rates result in considerable heating of tree bark surfaces¹³. Most species in closed canopy evergreen forests have thin bark³⁹, so fires effectively girdle and kill them. Larger, thicker-barked trees and others without vulnerable cambium layers (for example, palms) survive^{13,15,39}. In some forest types with root mats or deep organic layers, ground fires can accompany surface fires. These fires may consume everything down to the mineral soil and cause near-complete mortality^{15,35}. Fires in logged or previously burned forests are more severe, owing to greater fuel loads and lower humidities^{7,40,60}, with fireline intensities of $82\text{--}728\text{ kW m}^{-1}$ and more extreme fire behaviour¹³.

Although the chemistry and physics behind fire in tropical forests is no different from that in temperate or boreal forests, modelling fire susceptibility, behaviour, and effects in these forests will be more difficult owing to differences in tropical ecosystem characteristics. Closed-canopy tropical forests are species- and fuel-rich ecosystems that can regulate subcanopy humidity and hence the availability of fuels containing a variable mixture of numerous chemical compounds (for example, aliphatic and aromatic hydrocarbons,

alcohols, aldehydes, gums, sugars, terpenes, fats, waxes and oils)⁵². This ecosystem modulation of atmospheric humidity modifies fire behaviour in complex ways by changing the ignitability, sustainability and combustibility of fuels with varied structure and chemical composition.

Modelling of fire susceptibility in tropical evergreen forests is in its infancy. Fire susceptibility models must incorporate not only ambient weather conditions but also forest type, structure and disturbance history, as well as the availability of and ability to tap into deep soil moisture. Current conceptual models of susceptibility, the time since rain³⁹ and leaf-shedding^{65,66} fail to describe or provide the mechanistic understanding of forest moisture dynamics that are necessary to explain the behaviour of observed forest fires¹³. The deep-rooting nature⁴⁵ and tight hydrologic cycling⁴¹ of these forests make fire-susceptibility modelling challenging. Current models³² provide appropriate drought indices for these forests but cannot address the likelihood of actual fire ignition or spread owing to conceptual flaws that ignore the feedbacks of the forest upon the fire environment.

In the tropics, as elsewhere, models of fire spread and behaviour

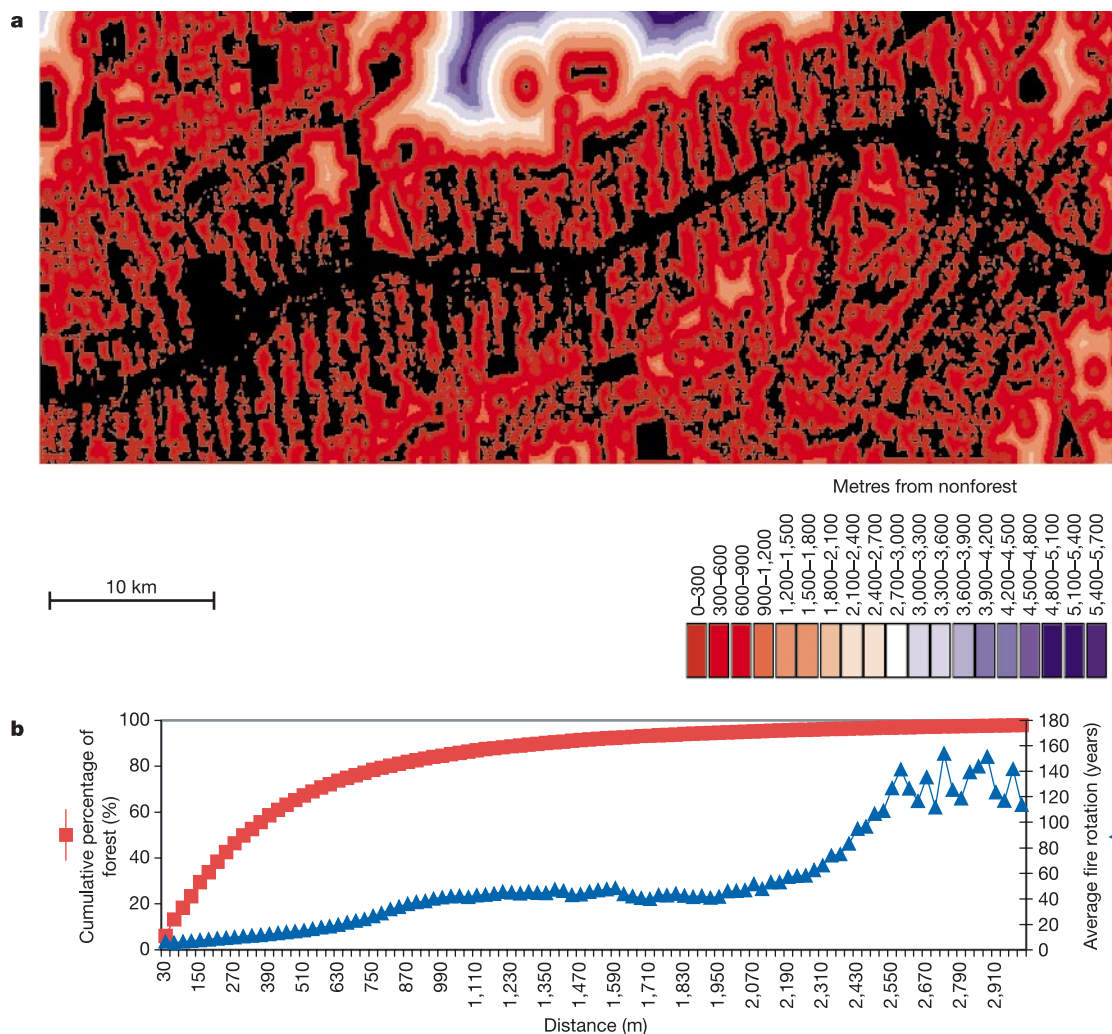


Figure 2 Spatial distribution of fire regimes⁶². **a**, Data is shown for Tailândia, Brazil, in 1997. Black areas are previously deforested, while other colours represent standing forests suffering different levels of fire impact. Forests in red are burning too frequently to persist as tropical evergreen forests and are turning into grassland and scrub¹³ in a continual erosion process^{43,63}. **b**, The graph shows both cumulative percentage of

remaining forests and fire frequency (average return interval in years) as a function of distance from deforested edges in the eastern Amazon. Tropical forests around the world are affected similarly but to different degrees depending on climate, fire load, and previous land cover changes or disturbances.

will require appropriate fuel models. Such fuel models have been standardized for many ecosystems and forest types⁶⁷ but not for tropical forests. Fires in tropical forests frequently occur in surface fuels, so understanding the characteristics and compositions of litter systems will be essential for predicting burning characteristics. Litter arrangement, loading, moisture, type and composition are all important modifiers of fire behaviour. Although there are a few quantitative descriptions of fuel quantities and size distributions^{13,39,68}, there do not appear to be comparable studies of fuel composition and chemistry. Litter production and decay rates vary widely in the tropics. Furthermore, litter from different species have large variations in energy values and flammability. Therefore, the phytosociological mix of litter-generating species can influence the flammability, sustainability, spread and heat-release characteristics of fires⁵². The fuel structures of forests with deep root mats or peat layers must also be understood because locally devastating ground fires in these substrates can have globally significant impacts¹⁸. With their extremely species-rich complements and varieties of canopy, surface and subsurface fuel structures, tropical forests will probably require a whole new set of fuel models to parameterize existing fire models such as BEHAVE Plus (<http://fire.org/cgi-bin/nav.cgi?pages=behave&mode=8>) and FARSITE⁶⁹. Dynamic relationships between fuels and forest moisture conditions will also be needed to address rapidly changing fuel availability and the potential involvement of green fuels.

Fire severity can be subjective but is often confused with intensity. This is understandable because severe fires in temperate and boreal forests are often intense surface or crown fires. The majority of tropical fires, however, are neither fast-spreading nor intense. They

are severe nevertheless because the impact to the forest is quite large^{7,8,10,15,40,60}. A severe fire can be rapidly spreading if it is intense but can also be of low intensity if it is slow-spreading. Relative severity for an ecosystem will depend on its specific adaptations, fortuitous or evolved, to the fires that are experienced. Prescribed burning is neither effective nor useful in tropical rainforests. Although burning favours the establishment of some commercial timber species⁷⁰ and temporarily reduces growth-inhibiting vine loads⁷¹, collateral damage from fire use outweighs any benefits^{70,71}. Fires are not useful for managing fuels either, because even very-low-intensity fires cause high mortality, creating more fuels than they consume¹³. Specific fire-protection and suppression recommendations are needed that can provide measurable results in reducing fire frequency.

Fires in tropical rainforests range from easily extinguished litter fires to nearly impossible-to-extinguish ground fires. In recurrent wildfires¹³, average fire intensities are ten times greater (30 versus 307 kW m⁻¹) and can spread twice as fast (0.25 versus 0.52 m min⁻¹). These limited data should not be considered to limit the range of possible fire conditions in tropical forests. In particular, the potential for landscape-level interactions among multiple fires to create large-scale, rapidly spreading conflagrations must be considered. Mass fires, also known as ‘firestorms’, can spread rapidly and lead to fire-induced weather, including high-velocity winds, lightning, and extreme fire behaviour (for example, fire whirls)⁷². There is historical precedence⁷³ for widespread logging damage, previous fire damage and widespread clearing fires combining to create mass fires in other forests, with devastating environmental, economic and human consequences. Regions with

Box 3

Fuels and moisture

Fuels are commonly divided into different size classes that have been determined to have specific moisture flux characteristics, specifically 1-h, 10-h, 100-h and 1,000-h fuel classes (0–0.62, 0.62–2.54, 2.54–7.62 and >7.62 cm). The number reflects the amount of time a fuel particle of a given size requires to reach 63% of equilibrium after a change in ambient moisture conditions⁹⁷. Smaller fuels (1-h) can either dry or become moist relatively quickly while larger fuels (1,000-h) take a long time to dry or recover moisture. Fuels ‘available’ for a given fire are simply those substances that require less heat to reach the point of combustion than is being radiated by the surrounding environment. Moisture increases the amount of heat necessary to reach the point of ignition. Sustained combustion occurs when ignited fuels conduct sufficient energy into adjacent fuels to start combustion before the currently burning fuels are consumed. If there is a continuous layer of fuels where this process can occur, then the fire will spread.

Closed-canopy forests are protected from the wind, and fire intensities and spread rates in such forests are also kept low by the high moisture content of fuels. But even at low intensities, slow-moving fires are severe because of the mortality caused by the long fire-contact times and the resultant heat transfer at the base of trees¹³. The moisture that protects tropical forests from fire is therefore also one of the reasons for the high mortality rates experienced when these forests do dry out enough to catch fire.

Owing to standard fuel monitoring practices in temperate forests, 10-h fuel-sticks have been used to evaluate fuel moisture and fire susceptibility in tropical forests^{35,39,60}. However, in many tropical forests, 1-h fuels carry the fires while 100-h and 1,000-h fuels contain smouldering fires throughout moister periods (for example, nights, short rains and so on), allowing fires to persist and spread¹³. Consequently, moisture levels in the 10-h fuel class may actually be the least appropriate for characterizing fire susceptibility in tropical forests.

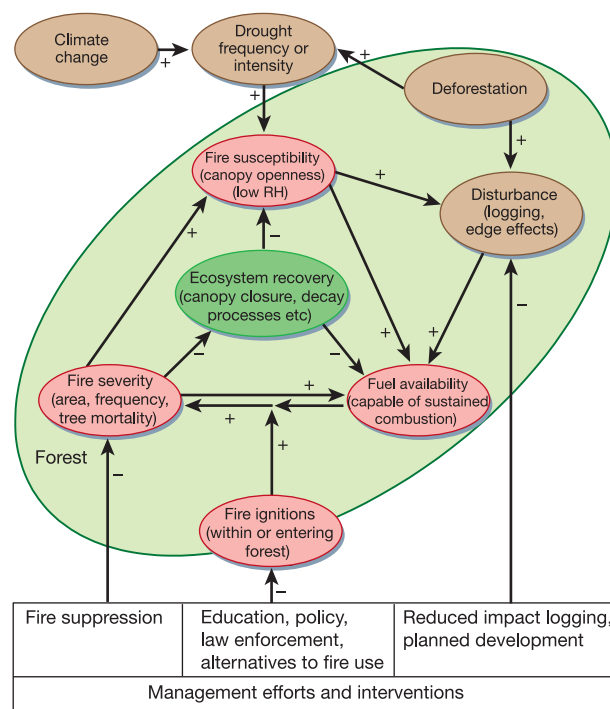


Figure 3 Positive and negative feedbacks controlling fire processes in tropical forests. Positioning indicates whether they occur within or outside the forest (green shading). Items in red control fire occurrence or behaviour. Items in brown modify the potential fire environment. Green indicates ecosystem processes acting in opposition to fire, specifically regrowth, canopy closure and decay of fuels. The management efforts and interventions box indicates how and where human actions can diminish tropical forest fires. Climate change encompasses effects of increased CO₂, land cover change, and aerosol loading that result in regional drying in the tropics^{26,27,79}. RH, relative humidity.

extensive logging- and fire-damaged forests that experience widespread and heavy fire use (such as East Kalimantan, Indonesia^{1,15,61}) are at greatest risk for such catastrophic fires.

Environmental impacts and climate feedbacks

Forest fires, deforestation and other forest disturbances effectively thin forests, reducing the amount of vegetation transpiring water. Reduced transpiration results in lowered local atmospheric humidity levels and increases the probability of future forest-fire occurrence⁷⁴. Decreased atmospheric moisture can negatively affect montane cloud forests by reducing the extent of and increasing the minimum altitude of downwind cloud fields⁷⁵. Regionally, transpiration from tropical forests is important for downwind precipitation, contributing 25% or more of annual rainfall in some regions^{43,44}.

Tropical forest fires reduce the ability of affected forests to retain water, exacerbating flooding, erosion and seasonal water shortages². Smoke-borne aerosols from fires disrupt normal hydrological processes and reduce rainfall^{28,29}, potentially contributing to regional drought. Through the radiative effects of smoke-derived aerosols, fires reduce the relative humidity and temperature gradients of the lower troposphere²⁷, disrupting normal cleansing processes and increasing average residence times of smoke particles in the atmosphere²⁶. Smoke plumes from tropical fires also increase the number and intensity of positive cloud-to-ground lightning strikes, long after and thousands of kilometres away from fires, altering 'natural' fire processes within and outside the tropics⁷⁶.

After a fire, leaves from newly killed trees quickly reblanket the ground while simultaneously opening the canopy, resulting in a more easily dried and fuel-laden forest^{7,13}. Whether burned^{7,14}, logged⁶⁰ or subjected to wind-throw and other edge effects^{48,49}, these forests become fire-susceptible after a few weeks without rain instead of several months⁷. Increased fire intensity and heat-transfer capacity of recurrent fires make them deadly to even larger thicker-barked trees¹³. Forest burning can create a positive feedback whereby more frequent and severe fires result in complete deforestation¹³ (Fig. 3).

Understanding fire in tropical rainforests

The existence of tropical forest fires has been periodically rediscovered and forgotten over the last century. Since Indonesia's large 1983 forest fires, studies of fire potential and effects have rapidly accrued in the literature but it is disturbing that Aubréville⁴³ was able to report much of this knowledge when he described fire effects in African forests in 1947. To move forward, tropical fire science needs to go beyond observation and description to provide increased understanding of the interactions between fire and tropical forests so that effective policy, management, modelling and monitoring efforts can be created.

Landscape-level knowledge of fire occurrence, behaviour and history is necessary for managing fire in the tropics. Tropical landscapes are a dynamic collection of intact, cleared, damaged and regenerating forests. To do more than chart the frequency and timing of burning at the landscape level, it will be necessary to relate fire-detection information to up-to-date land-cover maps. For ecological understanding of the importance of fire it will be necessary to know what is burning, how often and when. The need for high sampling frequency using high-resolution imagery currently makes the mapping of forest fires in the tropics expensive and labour-intensive². Where possible, such work should be combined with ongoing land use and land cover classifications to maximize the efficiency of resource use and the synergy of site information (for example, deforestation, selective logging, previous burning, forest regrowth and so on).

Fires have an impact on millions of hectares of rainforest², increase deforestation rates by as up to 50% in some regions¹³ and cause billions of dollars in damages in the tropics^{1,2}. To avoid the

continuous and insidious erosion⁴³ of forests by fire throughout the tropics, an increased level of attention, study and resources need to be directed to tropical fire science. This cannot be done by a blanket application of current fire knowledge from other parts of the world to the tropics, but neither does fire science need to be reinvented anew in tropical guise. Instead, current fire knowledge needs to be interpreted in the context of tropical forests and, where necessary, added to, by defining the mechanisms by which fire and ecosystem processes interact⁷⁷ in these forests. □

doi:10.1038/nature01437.

- Barber, C. V. & Schweithelm, J. *Trial by Fire: Forest Fire and Forestry Policy in Indonesia's Era of Crisis and Reform* (World Resources Institute, Washington DC, 2000).
- Cochrane, M. A. *Spreading Like Wildfire—Tropical Forest Fires in Latin America and the Caribbean: Prevention, Assessment and Early Warning* (United Nations Environment Programme (UNEP), Mexico City, 2002); available at http://www.rolac.unep.mx/dewalac/eng/fire_ingles.pdf.
- Ruitenbeek, J. in *Indonesia's Fire and Haze: the Cost of Catastrophe* (eds Glover, D. & Jessup, T.) 88–112 (International Development Research Centre, Ottawa, 1999).
- United Nations Environment Programme *GEO Latin America and the Caribbean: Environment Outlook 2000* (UNEP, Mexico City, 2000).
- Barbosa, R. I. & Fearnside, P. M. Incendios na Amazonia Brasileira: estimativa da emissão de gases do efeito estufa pela queima de diferentes ecossistemas de Roraima na passagem do evento El Niño (1997/98). *Acta Amazon.* **29**, 513–534 (1999).
- Goldammer, J. G. *et al.* in *Fire in the Tropical Biota* (ed. Goldammer, J. G.) 487–489 (Springer, Berlin, 1990).
- Cochrane, M. A. & Schulze, M. D. Fire as a recurrent event in tropical forests of the eastern Amazon: effects on forest structure, biomass, and species composition. *Biotropica* **31**, 2–16 (1999).
- Gerwing, J. J. Degradation of forests through logging and fire in the eastern Brazilian Amazon. *For. Ecol. Mgmt.* **157**, 131–141 (2002).
- Slik, J. W., Verburg, R. W. & Keblor, P. J. A. Effects of fire and selective logging on the tree species composition of lowland dipterocarp forest in East Kalimantan, Indonesia. *Biodiv. Conserv.* **11**, 85–98 (2002).
- Van Nieuwstadt, M. G. L., Sheil, D. & Kartawinata, K. The ecological consequences of logging in the burned forests of East Kalimantan, Indonesia. *Conserv. Biol.* **15**, 1183–1186 (2001).
- Kinnaird, M. F. & O'Brien, T. G. Ecological effects of wildfire on lowland rainforest in Sumatra. *Conserv. Biol.* **12**, 954–956 (1998).
- Kellman, M. & Meave, J. Fire in the tropical gallery forests of Belize. *J. Biogeogr.* **24**, 23–24 (1997).
- Cochrane, M. A. *et al.* Positive feedbacks in the fire dynamic of closed canopy tropical forests. *Science* **284**, 1832–1835 (1999).
- Swaine, M. D. Characteristics of dry forest in West Africa and the influence of fire. *J. Veg. Sci.* **3**, 365–374 (1992).
- Woods, P. Effects of logging, drought, and fire on structure and composition of tropical forests in Sabah, Malaysia. *Biotropica* **21**, 290–298 (1989).
- Fredericksen, N. J. & Fredericksen, T. S. Terrestrial wildlife responses to logging and fire in a Bolivian tropical humid forest. *Biodiv. Conserv.* **11**, 27–38 (2002).
- Barlow, J., Haugaasen, T. & Peres, C. A. Effects of ground fires on understory bird assemblages in Amazonian forests. *Biol. Conserv.* **105**, 157–169 (2002).
- Page, S. S., Siegfert, F., Rieley, J. O., Boehm, H. V. & Jaya, A. The amount of carbon released from peat and forest fires in Indonesia during 1997. *Nature* **420**, 61–65 (2003).
- Cairns, M. A., Hao, W. M., Alvarado, E. & Haggerty, P. in Vol. 1: *Proc. Joint Fire Science Conf. and Workshop: Crossing the Millennium: Integrating Spatial Technologies and Ecological Principles for a New Age in Fire Management* (eds Neuenchwander, L. F., Ryan, K. C., Gollberg, G. E. & Greer, J. D.) 242–247 (University of Idaho and the International Association of Wildland Fire, Moscow, Idaho, 2000).
- Phulpin, T., Lavenue, F., Bellan, M. F., Mougnot, B. & Blasco, F. Using SPOT-4 HRVIR and VEGETATION sensors to assess impact of tropical forest fires in Roraima, Brazil. *Int. J. Remote Sens.* **23**, 1943–1966 (2002).
- Houghton, J. T. (ed.) *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, UK, 2001).
- Kunii, O. in *Health Guidelines for Vegetation Fire Events: Background Papers* (eds Goh, K., Schwela, D., Goldammer, J. G. & Simpson, O.) 299–316 (World Health Organization, Geneva, 1999).
- Linden, E. A. Estrada do desastre. *Time Latina* (20 September 2000); available at <http://cnbrasil.com/2000/time/09/20/amazon/>.
- Ward, D. E. in *Health Guidelines for Vegetation Fire Events: Background Papers* (eds Goh, K., Schwela, D., Goldammer, J. G. & Simpson, O.) 71–86 (World Health Organization, Geneva, 1999).
- Crutzen, P. J. & Andreae, M. O. Biomass burning in the tropics: impact on atmospheric chemistry and biogeochemical cycles. *Science* **250**, 1669–1677 (1990).
- Ramanathan, V., Crutzen, P. J., Kiehl, J. T. & Rosenfeld, D. Aerosols, climate and the hydrological cycle. *Science* **294**, 2119–2124 (2001).
- Kaufman, Y. J., Tanré, D. & Boucher, O. A satellite view of aerosols in the climate system. *Nature* **419**, 215–223 (2002).
- Rosenfeld, D. TRMM observed first direct evidence of smoke from forest fires inhibiting rainfall. *Geophys. Res. Lett.* **26**, 3105–3108 (1999).
- Ackerman, A. S. *et al.* Reduction of tropical cloudiness by soot. *Science* **288**, 1024–1047 (2000).
- Pyne, S. J. *World Fire: The Culture of Fire on Earth* (Univ. Washington Press, Seattle, Washington, 1997).
- Meggors, B. J. Archeological evidence for the impact of Mega-Niño events on Amazonia during the past two millennia. *Clim. Change* **28**, 321–338 (1994).
- Nepstad, D. C., Moreira, A. G. & Alencar, A. A. *Flames in the Rain Forest: Origins, Impacts, and Alternatives to Amazonian Fires* (World Bank, Brasília, Brazil, 1999).
- Tate, G. H. H. Life zones at Mount Roraima. *Ecology* **13**, 235–257 (1932).

34. Goldammer, J. G. & Seibert, B. in *Fire in the Tropical Biota* (ed. Goldammer, J. G.) 11–31 (Springer, Berlin, 1990).
35. Uhl, C., Kauffman, J. B. & Cummings, D. L. Fire in the Venezuelan Amazon 2: Environmental conditions necessary for forest fires in the evergreen rainforest of Venezuela. *Oikos* **53**, 176–184 (1988).
36. Sanford, R. L., Saldarriaga, J., Clark, K., Uhl, C. & Herrera, R. Amazon rainforest fires. *Science* **227**, 53–55 (1985).
37. Goldammer, J. G. & Seibert, B. Natural rain-forest fires in Eastern Borneo during Pleistocene and Holocene. *Naturwissenschaften* **76**, 518–520 (1989).
38. Hammond, D. S. & ter Steege, H. Propensity for fire in Guianan rainforests. *Conserv. Biol.* **12**, 944–947 (1998).
39. Kauffman, J. B. & Uhl, C. in *Fire in the Tropical Biota* (ed. Goldammer, J. G.) 117–134 (Springer, Berlin, 1990).
40. Uhl, C. & Kauffman, J. B. Deforestation, fire susceptibility, and potential tree responses to fire in the eastern Amazon. *Ecology* **71**, 437–449 (1990).
41. Mueller-Dombois, M. *Proc. Conf. Fire Regimes and Ecosystem Properties*. General Technical Report WO-26, 137–176 (US Forest Service, Washington DC, 1981).
42. Moreira, M. Z. *et al.* Contribution of transpiration to forest ambient vapour based on isotopic measurements. *Glob. Change Biol.* **3**, 439–450 (1997).
43. Aubréville, A. M. A. The disappearance of the tropical forests of Africa. *Unasylva* **1**, 5–11 (1947).
44. Eltahir, E. A. B. & Bras, R. L. Precipitation recycling. *Rev. Geophys.* **34**, 367–378 (1996).
45. Nepstad, D. C. *et al.* The role of deep roots in water and carbon cycles of Amazonian forests and pastures. *Nature* **372**, 666–669 (1994).
46. Kleidon, A. & Lorenz, S. Deep roots sustain Amazonian rainforest in climate model simulations of the last ice age. *Geophys. Res. Lett.* **28**, 2425–2428 (2001).
47. Uhl, C. & Buschbacher, R. A disturbing synergism between cattle ranch burning practices and selective tree harvesting in the Eastern Amazon. *Biotropica* **17**, 265–268 (1985).
48. Cochrane, M. A. Synergistic interactions between habitat fragmentation and fire in evergreen tropical forests. *Conserv. Biol.* **15**, 1515–1521 (2001).
49. Cochrane, M. A. & Laurance, W. F. Fire as a large-scale edge effect in Amazonian forests. *J. Trop. Ecol.* **18**, 311–325 (2002).
50. Cochrane, M. A., Skole, D. L., Matricardi, E. A. T., Barber, C., Chomentowski, W. in *Working Forests in the Tropics: Conservation through Sustainable Management?* (eds Zarin, D. J. *et al.*) (Columbia Univ. Press, New York, in the press).
51. Bucini, G. & Lambin, E. F. Fire impacts on vegetation in Central Africa: a remote-sensing-based statistical analysis. *Appl. Geogr.* **22**, 27–48 (2002).
52. Stott, P. Combustion in tropical biomass fires: a critical review. *Prog. Phys. Geogr.* **24**, 355–377 (2000).
53. Tutin, C. E. G., White, L. J. T. & Mackangamissandzou, A. Lightning strike burns large forest tree in the Lopé Reserve, Gabon. *Glob. Ecol. Biogeogr. Lett.* **5**, 36–41 (1996).
54. Vayda, A. P. *Finding Causes of the 1997–98 Indonesian Forest Fires: Problems and Possibilities* (World Wide Fund for Nature (WWF), Jakarta, Indonesia, 1999).
55. Skole, D. & Tucker, C. J. Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988. *Science* **260**, 1905–1910 (1993).
56. Laurance, W. F. *et al.* The future of the Brazilian Amazon: Development trends and deforestation. *Science* **291**, 438–439 (2001).
57. Laurance, W. F. *et al.* Biomass collapse in Amazonian forest fragments. *Science* **278**, 1117–1118 (1997).
58. Nascimento, H. E. M. & Laurance, W. F. Total aboveground biomass in central Amazonian rainforests: a landscape-scale study. *For. Ecol. Mgmt* **168**, 311–321 (2002).
59. Verissimo, A., Barreto, P., Tarifa, R. & Uhl, C. Extraction of a high-value natural source from Amazon: the case of mahogany. *For. Ecol. Mgmt* **72**, 39–60 (1995).
60. Holdsworth, A. R. & Uhl, C. Fire in Amazonian selectively logged rain forest and the potential for fire reduction. *Ecol. Appl.* **7**, 713–725 (1997).
61. Siegert, F., Ruecker, G., Hinrichs, A. & Hoffman, A. A. Increased damage from fires in logged forests during droughts caused by El Niño. *Nature* **414**, 437–440 (2001).
62. Cochrane, M. A. In the line of fire: Understanding the impacts of tropical forest fires. *Environment* **43**, 28–38 (2001).
63. Gascon, C., Williamson, G. B. & Fonseca, G. A. B. Receding edges and vanishing fragments. *Science* **288**, 1356–1358 (2000).
64. Siegert, F. & Ruecker, G. Use of multitemporal ERS-2 SAR images for identification of burned scars in south-east Asian tropical rainforest. *Int. J. Remote Sens.* **21**, 831–837 (2000).
65. Huber, O., Steyermark, J. A., Prance, G. T. & Alés, C. The vegetation of the Sierra Parima, Venezuela-Brazil: Some results of recent exploration. *Brittonia* **36**, 104–139 (1984).
66. Nepstad, D. C., Jipp, P., Moutinho, P., Negreiros, G. & Vieira, S. in *Evaluating and Monitoring the Health of Large-Scale Ecosystems* (eds Rapport, D., Gander, C. & Calow, P.) 333–349 (Springer, New York, 1995).
67. Anderson, H. E. *Aids to Determining Fuel Models for Estimating Fire Behavior*. General Technical Report INT-122 (US Forest Service, Ogden, Utah, 1982).
68. Kauffman, J. B., Cummings, D. L. & Ward, D. E. Fire in the Brazilian Amazon 2. Biomass, nutrient pools and losses in cattle pastures. *Oecologia* **113**, 415–427 (1998).
69. Finney, M. A. *FARSITE: Fire Area Simulator—Model Development and Evaluation* (US Forest Service, Ogden, Utah, 1998).
70. Gould, K. A. *et al.* Post-fire tree regeneration in lowland Bolivia: implications for fire management. *For. Ecol. Mgmt* **165**, 225–234 (2002).
71. Gerwing, J. J. Testing liana cutting and controlled burning as silvicultural treatments for a logged forest in the eastern Amazon. *J. Appl. Ecol.* **38**, 1264–1276 (2001).
72. Quintero, J. G. Canadian mass fire experiment. 1989. *J. Fire Protect. Eng.* **5**, 67–78 (1993).
73. Bailey, W. O. *Report on the Michigan Forest Fires of 1881* (US War Department, Washington DC, 1882).
74. Laurance, W. F. & Williamson, B. Positive feedbacks among forest fragmentation, drought, and climate change in the Amazon. *Conserv. Biol.* **15**, 1529–1535 (2001).
75. Lawton, R. O., Nair, U. S., Pielke, R. A. & Welch, R. M. Climatic impact of tropical lowland deforestation on nearby montane cloud forests. *Science* **294**, 584–587 (2001).
76. Lyons, W. A., Nelson, T. E., Williams, E. R., Cramer, J. A. & Turner, T. R. Enhanced positive cloud-to-ground lightning in thunderstorms ingesting smoke from fires. *Science* **282**, 77–80 (1998).
77. Johnson, E. A. & Miyanishi, K. *Forest Fires: Behavior and Ecological Effects* (Academic, San Diego, California, 2001).
78. Verissimo, A., Cochrane, M. A. & Souza, C. Jr National forests in the Amazon. *Science* **297**, 1478 (2002).
79. DeFries, R. S., Bounoua, L. & Collatz, G. J. Human modification of the landscape and surface climate in the next fifty years. *Glob. Change Biol.* **8**, 438–458 (2002).
80. Giglio, L., Kendall, J. D. & Justice, C. O. Evaluation of global fire detection algorithms using simulated AVHRR infrared data. *Int. J. Remote Sens.* **20**, 1947–1985 (1999).
81. Elvidge, C. D. *et al.* in *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications* (eds Lunetta, R. S. & Elvidge, C. D.) 74–122 (Ann Arbor Press, Ann Arbor, Michigan, 1999).
82. Fuller, D. O. & Fulk, M. Comparison of NOAA-AVHRR and DMSP-OLS for operational fire monitoring in Kalimantan, Indonesia. *Int. J. Remote Sens.* **21**, 181–187 (2000).
83. Pereira, A. C. & Setzer, A. W. Comparison of fire detection in savannas using AVHRR's channel 3 and TM images. *Int. J. Remote Sens.* **17**, 1925–1937 (1996).
84. Fuller, D. O. Satellite remote sensing of biomass burning with optical and thermal sensors. *Prog. Phys. Geogr.* **44**, 543–561 (2000).
85. Giglio, L., Kendall, J. D. & Tucker, C. J. Remote sensing of fires with the TRMM VIRS. *Int. J. Remote Sens.* **21**, 203–207 (2000).
86. Goldammer, J. G. in *Health Guidelines for Vegetation Fire Events: Background Papers* (eds Goh, K., Schwela, D., Goldammer, J. G. & Simpson, O.) 9–70 (World Health Organization, Geneva, 1999).
87. Oertel, D. *et al.* in *Forest Fire Monitoring and Mapping: a Component of Global Observation of Forest Cover* (eds Ahern, F., Grégoire, J. M. & Justice, C.) 224–228 (Joint Research Centre, Ispra, Italy, 2000).
88. Eva, H. & Lambin, E. Burnt area mapping in Central Africa using ATSR data. *Int. J. Remote Sens.* **19**, 3473–3497 (1998).
89. Cochrane, M. A. & Souza, C. M. Jr Linear mixture model classification of burned forests in the eastern Amazon. *Int. J. Remote Sens.* **19**, 3433–3440 (1998).
90. Stone, T. & Lefebvre, P. Using multi-temporal satellite data to evaluate selective logging in Pará. *Brazil. Int. J. Remote Sens.* **19**, 2517–2526 (1998).
91. Nepstad, D. C. *et al.* Large-scale impoverishment of Amazonian forests by logging and fire. *Nature* **398**, 505–508 (1999).
92. Matricardi, E. A. T., Skole, D. L., Chomentowski, W. & Cochrane, M. A. *Multi-Temporal Detection and Measurement of Selective Logging in the Brazilian Amazon Using Remote Sensing* (Michigan State Univ., East Lansing, Michigan, 2001); available at <http://www.globalchange.msu.edu/publications>.
93. Uhl, C. *et al.* Natural resource management in the Brazilian Amazon. *Bioscience* **47**, 160–168 (1997).
94. Jackson, S. M., Fredericksen, T. S. & Malcolm, J. R. Area disturbed and residual stand damage following logging in a Bolivian tropical forest. *For. Ecol. Mgmt* **166**, 271–283 (2002).
95. Sist, P. & Nguyen-Thé, N. Logging damage and the subsequent dynamics of a dipterocarp forest in East Kalimantan. *For. Ecol. Mgmt* **165**, 85–103 (2002).
96. Goldammer, J. G., Seibert, B. & Schindele, W. in *Dipterocarp Forest Ecosystems: Towards Sustainable Management* (eds Schulte, A. & Schöne, D.) 155–185 (World Scientific, Singapore, 1996).
97. Agee, J. K. *Fire Ecology of Pacific Northwest Forests* (Island Press, Washington DC, 1993).

Acknowledgements I am grateful for support from NASA.

Correspondence and requests for materials should be addressed to the author (e-mail: cochrane@globalchange.msu.edu).

Energetic neutral atoms from a trans-Europa gas torus at Jupiter

B. H. Mauk, D. G. Mitchell, S. M. Krimigis, E. C. Roelof & C. P. Paranicas

The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA

The space environments—or magnetospheres—of magnetized planets emit copious quantities of energetic neutral atoms (ENAs) at energies between tens of electron volts to hundreds of kiloelectron volts (keV)^{1,2}. These energetic atoms result from charge exchange between magnetically trapped energetic ions and cold neutral atoms, and they carry significant amounts of energy and mass from the magnetospheres. Imaging their distribution allows us to investigate the structure of planetary magnetospheres^{3–6}. Here we report the analysis of 50–80 keV ENA images of Jupiter's magnetosphere⁷, where two distinct emission regions dominate: the upper atmosphere of Jupiter itself, and a torus of emission residing just outside the orbit of Jupiter's satellite Europa. The trans-Europa component shows that, unexpectedly, Europa generates a gas cloud comparable in gas content to that associated with the volcanic moon Io. The quantity of gas found indicates that Europa has a much greater impact than hitherto believed on the structure of, and the energy flow within, Jupiter's magnetosphere.

At Earth trapped ions are converted to ENAs when they interact with cold neutral hydrogen atoms that constitute Earth's extended upper atmosphere, its exosphere⁸. The gas density falls roughly exponentially with altitude, but remains at values useful for probing the magnetosphere to altitudes greater than ten Earth radii⁹. At giant planets such as Jupiter, exospheric neutrals are much more tightly bound to the planet by gravity. However, at Jupiter it is known that the volcanic moon Io, orbiting at 5.9 Jupiter radii, is an abundant source of neutral gas principally in the form of volcanically spewed sulphur dioxide (SO₂). When a small fraction of the total gas volume is liberated from Io by interactions with Jupiter's space environment, it dissociates into torus-like clouds of mostly S and O atoms that encircle Jupiter^{10,11}. When ENAs were first detected coming from Jupiter in 1979 by the Voyager spacecraft^{12,13}, it was assumed that Io provided the gases needed to convert the trapped energetic ions. The gases from Io are expected to disperse sparsely yet broadly throughout Jupiter's magnetosphere^{13,14}, so that densities high enough to be useful for performing ENA imaging were thought to extend to distances from Jupiter as large as 30 Jupiter radii¹⁵.

An ENA imager was included with the Cassini mission to orbit Saturn^{16,17}. This ion and neutral camera (INCA) was operated during a fly-by of Jupiter in late 2000 and early 2001 (ref. 7). However, the INCA sensor was designed to resolve magnetospheric sources of ENAs from distances of a few to at most tens of planetary radii. The closest that Cassini came to Jupiter was around 140 Jupiter radii. Thus, it is not possible to ascertain the detailed Jupiter ENA source regions using the raw ENA images (Fig. 1a)⁷. The Jupiter ENA images have now been processed to remove instrumental distortions using a point spread function (PSF) derived using images of Jupiter taken from distances greater than 800 Jupiter radii (Fig. 1b). The PSF is used in standard mathematical procedures¹⁸ to sharpen the images obtained at around 140 Jupiter radii.

The processed images reveal interesting features (Fig. 1c). The 'trans-Europa' torus component reveals itself as the two outermost dot-like features, each centred just outside the orbit of Europa (orbiting at ~9.5 Jupiter radii). The features are dot-like because of limb brightening, a consequence of the viewing geometry. The torus

is viewed edge-on, and the image is brightest at line-of-sight angles that pass through the greatest volume of the ENA emission regions. The limb brightening is accentuated by the sensitivity of the ENA emission process to the angle α between the line-of-sight and the magnetic field lines that trap the ions. The total emission rate of

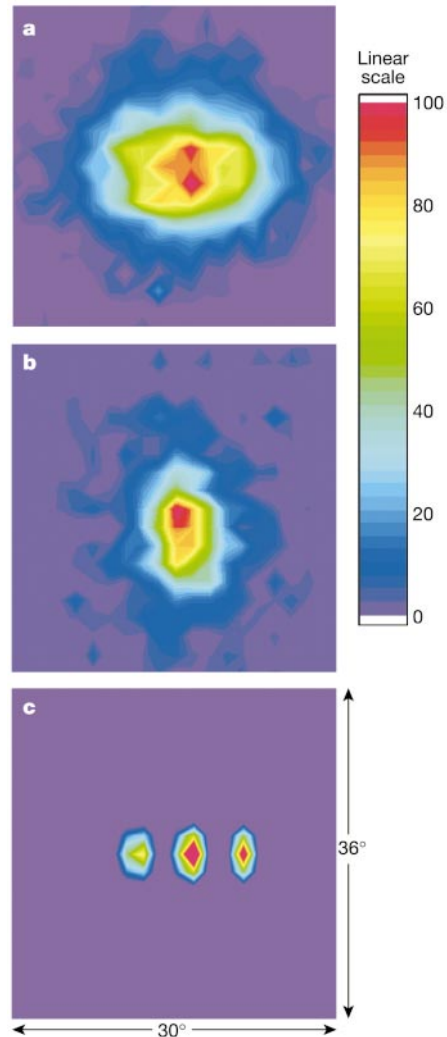


Figure 1 Energetic neutral atom (ENA) images of Jupiter's magnetosphere. **a**, Raw (uncorrected) image taken from a distance of around 140 Jupiter radii⁷. The ion and neutral camera (INCA) sensor^{16,17} on the Cassini spacecraft took this 15-h image beginning at 2200 Universal Time on 1 January 2001. Red represents ~190 counts within $1.4^\circ \times 1.9^\circ$ pixels. The ENAs are hydrogen atoms with speeds between 3,000 and 4,000 km s⁻¹ (~50–80 keV) but with a fraction of heavy atoms (O or S) of the same speed⁷. Much of the spread results from sensor characteristics. **b**, Raw (uncorrected) image taken from >800 Jupiter radii. It combines ten accumulations of ~10 h per day during the Cassini approach to Jupiter. Red represents ~164 counts. From this distance Jupiter behaves as a point source for INCA, and thus the image represents the INCA point spread function (PSF). A reasonable representation is $PSF = \exp[-(x - x_0)^2 / (2\sigma_x^2)] \times \exp[-(y - y_0)^2 / (2\sigma_y^2)]$, where all 'x' units are ~1.4° pixels, all 'y' units are ~1.9° pixels, and $(\sigma_x, \sigma_y) = (2.3, 3.2)$, valid only for the 50–80 keV H⁺ channel. Translational invariance of the PSF is assured by ground calibration. However, the full measure of the PSF is determined with flight data because the INCA aperture could not be fully illuminated by ground accelerator beams. **c**, Deconvolved ENA image of Jupiter's magnetosphere from ~140 Jupiter radii. The raw image of **a** was processed (deconvolved) with the analytic representation of the INCA PSF (**b**) to reduce the instrumental distortions¹⁸. Red represents ~1,450 counts. A floor of 10% of the peak value suppresses processing artefacts (Fig. 2). The central feature is centred on Jupiter, and the outermost features reside just beyond the orbit of the satellite Europa, orbiting at ~9.5 Jupiter radii (Fig. 2).

trans-Europa, 50–80 keV ENAs is roughly $0.8 \times 10^{24} \text{ s}^{-1}$. The Jupiter atmospheric source, the middle feature centred on Jupiter, is identified as coming from Jupiter's upper atmosphere based on lessons learned with ENA images of Earth's magnetosphere^{1,3,5}. The total emission rate of the Jupiter atmospheric 50–80 keV ENAs is roughly $0.5 \times 10^{24} \text{ s}^{-1}$. One-dimensional versions of the raw and processed images (Fig. 2a) show large differences that engender concerns over the robustness and uniqueness of the image processing. Of the many procedures used to assure the robustness of the results, the most constraining is the comparison between the final image and the intensity of trapped, 50–80 keV protons in Jupiter's magnetosphere, as measured from the Jupiter-orbiting Galileo spacecraft (Fig. 2b)¹⁹.

The comparison between the trapped ion intensities and the ENA emission intensities (Fig. 2b) also quantifies the neutral gas needed to explain the ENA intensities. To estimate the total number of gas atoms N outside, but within 5 Jupiter radii of, Europa's orbit, the integral equation is used that relates the ENA intensities $I_{\text{ENA}}(\mathbf{s}, E)$ to the ion intensities $I_i(E)$, neutral gas densities n , and the charge-exchange cross-section σ , written $I_{\text{ENA}}(\mathbf{s}, E) = \int I_i(E) n \sigma d\mathbf{s}$, where $\mathbf{s}$

is distance along the line-of-sight direction, $\mathbf{s}$ is its unit vector, and E is energy. Because of the viewing geometry and the fact that the neutral gases are confined close to Jupiter's equator, the peak ENA emissions are insensitive to the α -parameter described in the preceding paragraph, and so that dependence is not shown here. Assuming H^+ on H (or H_2) interactions, we obtain $N \approx 4.5 \times 10^{33} \pm 20\%$ atoms or molecules, where the error range is associated with assuming a range of gas torus radial thicknesses, from 1 to 5 Jupiter radii (the 1 Jupiter radius assumption yields the high N estimate because the emission volume is less). These estimates are halved for H^+ on O interactions. For a trans-Europa emission region with both a radial extent and a height of 2 Jupiter radii, the gas density is about 40 atoms cm^{-3} . This number is comparable to a previous inference derived on the basis of the angular distributions of hot ions measured near Europa's orbit with Galileo (A. Lagg, N. Krupp, J. Woch and D. J. Williams, unpublished work). This density is a factor of 50 (ref. 13) to 100 (ref. 14) greater than that expected on the basis of the dispersal of neutral gases from Io. It is a substantial surprise that gases likely to have come from Europa predominate in the conversion of hot Jupiter ions to ENAs^{12–15}. Also, while the sources of energetic protons in Jupiter's magnetosphere have been attributed to either Jupiter's atmosphere or the wind of ionized gases from the Sun (the solar wind)^{20,21}, the Europa gas torus now becomes a prime suspect. Assuming radial symmetry about Europa's orbit (the preceding calculations considered only contributions outside of Europa's orbit), the total content of the Europa gas cloud is twice the value given above, or around 9×10^{33} atoms or molecules (probably composed predominantly of H , H_2 , O and O_2 from the water-ice surface of Europa^{22,23}). Surprisingly, this value is comparable to the highest of the varying estimates of the content of Io's gas torus, ranging from about 2×10^{33} (40 cm^{-3} over the volume of the plasma torus)¹⁴, down to an order of magnitude smaller¹³, and up to about 8×10^{33} (ref. 24).

The comparison between the Io and Europa gas tori becomes perhaps less surprising when one considers the large uncertainties in the mass loss rates from Europa and the content of the Io gas torus. Water-ice products are removed from Europa's surface and atmosphere, so it is thought, by sputtering that results from the impact of energetic and fast-flowing ions^{25,26}. Calculations based on estimates of satellite mass loss rates and on estimated partitioning into different atomic and molecular species^{22,23} yield values between $\sim 5 \times 10^{26}$ (ref. 25) and 4×10^{27} (ref. 26) atoms or molecules liberated per second. These values are 1 to 2 orders of magnitude less than the purported source of S and O from Io ($2 \times 10^{28} \text{ s}^{-1}$ (ref. 10)). Given the relative contents of the respective gas tori near Io and Europa, either the subsequent loss rates of Io torus gases are 1 to 2 orders of magnitude greater than the loss rates of Europa torus gases, or the source rate of gases from Europa has been underestimated. The composition and the plasma environment of the Io gas torus are substantially different from that of the Europa gas torus, and so the torus gas loss rates are undoubtedly different. However, there has been a suggestion that the sputtered source rates from Europa are amplified by as much as an order of magnitude by secondary particles that are accelerated by the strong electric fields generated by the satellite-magnetosphere interaction process²³. Although this claim has been disputed²⁶, it deserves a second look on the basis of the results reported here.

The Jupiter atmospheric source of ENAs (central feature in Fig. 1c) provides a unique diagnostic of the second way that trapped energetic ions are lost from the magnetospheric system. Trapped charged particles bounce back and forth between northern and southern magnetic mirror points as they travel along the dipole-shaped magnetic field lines. Scattering by waves that propagate within the magnetospheric medium cause both mirror points to slowly migrate to lower and lower altitudes, towards the atmosphere. On interacting with the atmosphere some ions escape as ENAs and some are captured by the atmosphere. If the ions

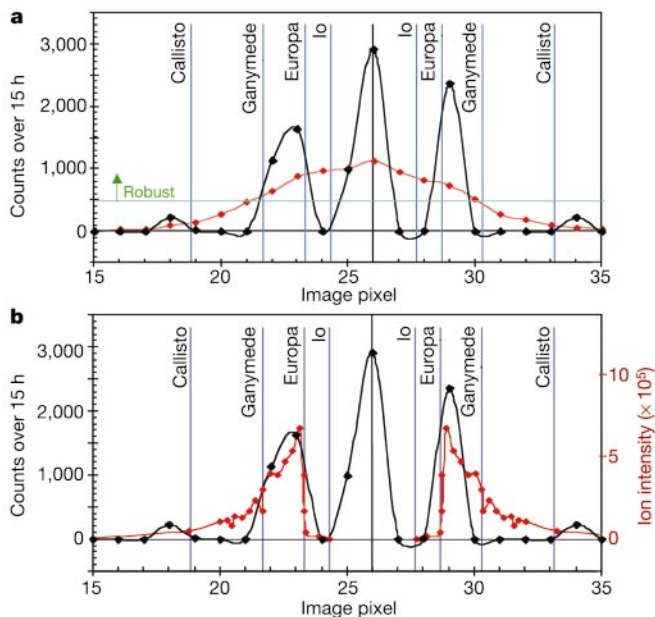


Figure 2 One-dimensional energetic neutral atom (ENA) images of Jupiter's magnetosphere. **a**, One-dimensional raw and corrected images of Jupiter from a distance of ~ 140 Jupiter radii. The profiles were derived by summing the vertical (y) pixels in the raw image of Fig. 1a (red curve) and the processed (deconvolved) image of Fig. 1c (black curve). The vertical blue lines are the tangent-line distances of the moons Callisto, Ganymede, Europa and Io, orbiting respectively at radial distances of ~ 26 , ~ 15 , ~ 9.5 and ~ 5.9 Jupiter radii. The horizontal green line demarcates a boundary between processed image features that are robust to perturbations to the deconvolution process (above the line) and those features that are not robust (below the line). Specifically, the 'bump' in the processed image just outside of Callisto is not real. **b**, Comparison between the one-dimensional processed ENA image of Jupiter (black curve) and the intensities of protons, energy-integrated from 50 to 80 keV, obtained *in situ* by the Jupiter-orbiting Galileo energetic particle detector¹⁹ (red curve in units of $\text{cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$). The energy-integrated oxygen and sulphur ions that can contribute to the INCA image (those with the same speed as the protons, and thus also with roughly the same charge exchange cross-sections) are a factor of about eight lower than the proton intensities throughout the region of interest. Multiple charging of S and O ions further diminishes heavy-ion contributions. The *in situ* proton intensities represent selective averages and samples of data obtained during eight different Galileo orbits. The sharpness of the discontinuity at Europa may be an exaggeration of the average configuration because different Galileo orbits were used to sample values inside and outside Europa (Europa's gravity was often used to redirect Galileo's trajectory).

originate between Europa and Ganymede (Fig. 2b), we obtain more than $1 \times 10^4 \text{ cm}^{-2} \text{ s}^{-1}$ for the atmospheric flux of ions that are converted to escaping 50–80 keV ENAs. This value establishes a lower limit to the precipitating intensities equal to 0.7% of the high-altitude, trapped ion intensities. If all protons measured by Galileo precipitate with a similar fraction, the energy deposition into the atmosphere would be of order $0.1\text{--}0.2 \text{ erg cm}^{-2} \text{ s}^{-1}$, at magnetic latitudes roughly between 71° and 75° (latitudes that map magnetically to the region between the orbits of Ganymede and Europa). Here it is assumed that the number of ions captured by the atmosphere is comparable to the number of escaping ENAs. Such values are several orders of magnitude smaller than the electron energy fluxes associated with Jupiter's main aurora²⁷, and are probably too low to stimulate presently observable northern or southern (auroral) light emissions.

The 50–80 keV ENA emissions documented here represent a small fraction of the total ENA emissions at all energies. If we assume very crudely that the proportion of trapped 50–80 keV protons with respect to the entire proton distribution as measured by Galileo in the trans-Europa region holds true for the ENA emission spectra, then the total ENA numbers and powers emitted from the trans-Europa region and Jupiter's atmosphere are around 10^{25} s^{-1} and 10^{12} W for each region. Although the estimated total numbers of emitted ENAs are much lower than the number of low-energy ($<0.6 \text{ keV}$) O and S neutrals emitted from the Io torus ($\sim 10^{28} \text{ s}^{-1}$) (ref. 10), the emitted powers are comparable. The total estimated power emitted for each of the ENA source regions is roughly an order of magnitude below the energy dissipation that occurs within the electron-stimulated main auroral regions of Jupiter²⁷. □

Received 12 November 2002; accepted 13 January 2003; doi:10.1038/nature01431.

1. Roelof, E. C. Energetic neutral atom image of a storm-time ring current. *Geophys. Res. Lett.* **14**, 652–655 (1987).
2. Mendillo, M., Baumgardner, J., Flynn, B. & Hughes, W. J. The extended sodium nebula of Jupiter. *Nature* **348**, 312–314 (1990).
3. Brandt, P. C., Barabash, S., Roelof, E. C. & Chase, C. J. Energetic neutral atom imaging at low altitudes from the Swedish microsatellite Astrid: Observations at low ($\leq 10 \text{ keV}$) energies. *J. Geophys. Res.* **106**, 24663–24674 (2001).
4. Henderson, M. G. *et al.* First energetic neutral atom images from POLAR. *Geophys. Res. Lett.* **24**, 1167–1170 (1997).
5. Mitchell, D. G. *et al.* Imaging two geomagnetic storms in energetic neutral atoms. *Geophys. Res. Lett.* **28**, 1151–1154 (2001).
6. Burch, J. L. *et al.* Views of Earth's magnetosphere with the IMAGE satellite. *Science* **291**, 619–624 (2001).
7. Krimigis, S. M. *et al.* A nebula of gases from Io surrounding Jupiter. *Nature* **415**, 994–996 (2002).
8. Rairden, R. L., Frank, L. A. & Craven, J. D. Geocoronal imaging with Dynamics Explorer. *J. Geophys. Res.* **91**, 13613–13630 (1986).
9. Brandt, P. C. *et al.* IMAGE/HENA: Global ENA imaging of the plasmasheet and ring current during substorms. *J. Geophys. Res.* (in the press).
10. Smyth, W. H. & Marconi, M. L. Io's oxygen source: Determination from ground-based observations and implications for the plasma torus. *J. Geophys. Res.* **105**, 7783–7792 (2000).
11. Brown, R. A. The Jupiter hot plasma torus: observed electron temperature and energy flows. *Astrophys. J.* **244**, 1072–1080 (1981).
12. Kirsch, E. *et al.* Upper limits for x-ray and energetic neutral particle emission from Jupiter: Voyager 1 results. *Geophys. Res. Lett.* **8**, 169–172 (1981).
13. Cheng, A. F. Energetic neutral particles from Jupiter and Saturn. *J. Geophys. Res.* **91**, 4524–4530 (1986).
14. Schreier, R., Eviatar, A. & Vasyliunas, V. M. A two-dimensional model of plasma transport and chemistry in the Jovian magnetosphere. *J. Geophys. Res.* **103**, 19901–19914 (1998).
15. Mauk, B. H. *et al.* Energetic neutral atom imaging of Jupiter's magnetosphere using the Cassini MIMI instrument. *Adv. Space Res.* **21**, 1483–1486 (1998).
16. Mitchell, D. G. *et al.* The imaging neutral camera (INCA) for the NASA Cassini mission to Saturn/Titan. *Proc. SPIE Int. Symp. Opt. Sci. Eng. Instrum. Mission to the Sun* **2803**, 154–161 (1996).
17. Krimigis, S. M. *et al.* The Magnetospheric Imaging Instrument (MIMI) on the Cassini mission to Saturn/Titan. *Space Sci. Rev.* (in the press).
18. Rosenfeld, A. & Kak, A. C. *Digital Picture Processing* (Academic Press, Inc., Orlando, Florida, 1982).
19. Williams, D. J. *et al.* Electron beams and ion composition measured at Io and in its torus. *Science* **274**, 401–403 (1996).
20. Bagenal, F. Giant planet magnetospheres. *Annu. Rev. Earth Planet. Sci.* **10**, 289–328 (1992).
21. Khurana, K. K. *et al.* in *Jupiter* (ed. Bagenal, F.) (Univ. Arizona Press, Tucson, in the press).
22. Bar-Nun, A. *et al.* Ejection of H_2O , O_2 , H_2 , and H from water ice by 0.5–6 keV H^+ and Ne^+ ion bombardment. *Surf. Sci.* **150**, 143–156 (1985).
23. Ip, W.-H. *et al.* Ion sputtering and surface erosion at Europa. *Geophys. Res. Lett.* **25**, 829–832 (1998).
24. Skinner, T. E. & Durrance, S. T. Neutral oxygen and sulfur densities in the Io torus. *Astrophys. J.* **310**, 966–971 (1986).

25. Cooper, J. F. *et al.* Energetic ion and electron irradiation of the icy Galilean satellites. *Icarus* **149**, 133–159 (2001).
26. Saur, J., Strobel, D. F. & Neubauer, F. M. Interaction of the Jovian magnetosphere with Europa: Constraints on the neutral atmosphere. *J. Geophys. Res.* **103**, 19947–19962 (1998).
27. Clarke, J. T. *et al.* Far ultraviolet imaging of Jupiter's aurora and the Io 'footprint'. *Science* **274**, 404–409 (1996).

Acknowledgements We gratefully acknowledge the support of NASA through the Jet Propulsion Laboratory for both the Cassini and Galileo Projects, which made this work possible.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.H.M. (e-mail: Barry.Mauk@jhuapl.edu).

Upper limits to submillimetre-range forces from extra space-time dimensions

Joshua C. Long^{*†}, Hilton W. Chan^{*†}, Allison B. Churnside^{*}, Eric A. Gulbis^{*}, Michael C. M. Varney^{*} & John C. Price^{*}

^{*} Physics Department, University of Colorado, UCB 390, Boulder, Colorado 80309, USA

String theory is the most promising approach to the long-sought unified description of the four forces of nature and the elementary particles¹, but direct evidence supporting it is lacking. The theory requires six extra spatial dimensions beyond the three that we observe; it is usually supposed that these extra dimensions are curled up into small spaces. This 'compactification' induces 'moduli' fields, which describe the size and shape of the compact dimensions at each point in space-time. These moduli fields generate forces with strengths comparable to gravity, which according to some recent predictions^{2–7} might be detected on length scales of about $100 \mu\text{m}$. Here we report a search for gravitational-strength forces using planar oscillators separated by a gap of $108 \mu\text{m}$. No new forces are observed, ruling out a substantial portion of the previously allowed parameter space⁴ for the strange and gluon moduli forces, and setting a new upper limit on the range of the string dilaton^{2,3} and radion^{5–7} forces.

The combined potential energy V due to a modulus force and newtonian gravity may be written:

$$V = - \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{G\rho_1(\mathbf{r}_1)\rho_2(\mathbf{r}_2)}{r_{12}} [1 + \alpha \exp(-r_{12}/\lambda)] \quad (1)$$

The first term is Newton's universal gravitation law, with G the gravitational constant, r_{12} the distance between two points $\mathbf{r}_1$ and $\mathbf{r}_2$ in the test masses, and ρ_1 , ρ_2 the mass densities of the two bodies. The second term is a Yukawa potential, with α the strength of the new force relative to gravity, and λ the range. As for any force-mediating field, the range of the modulus force is related to its mass m by $\lambda = \hbar/mc$. Previous tests of Newtonian gravity and searches for new macroscopic forces have covered length scales from light-years to nanometres^{8–10}, and it has been found that new forces of gravitational strength can be excluded for ranges λ from $200 \mu\text{m}$ to nearly a light-year^{8,11,12}, but limits on new forces become poor very rapidly below $200 \mu\text{m}$ (refs 9, 13).

[†] Present addresses: Los Alamos Neutron Science Center, LANSCE-3, MS-H855, Los Alamos, New Mexico 87545, USA (J.C.L.); and Physics Department, Stanford University, Stanford, California 94305, USA (H.W.C.).

The strange and gluon moduli arise in a scenario where supersymmetry (a hypothesized symmetry incorporated in string theory that relates bosons to fermions) is broken at 10–100 TeV, and all compactification occurs near the ultrahigh Planck scale of 10^{19} GeV where gravity is unified (in these scenarios) with other forces. In this case a range of approximately a millimetre is specifically predicted. The dilaton is a scalar field required for the consistency of string theory whose vacuum value fixes the strength of the interaction between strings. It may be interpreted as the modulus describing the size of the tenth space dimension that is compactified in M-theory to yield string theory in nine space dimensions. (M-theory is the structure that unifies the various types of string theory into a single framework¹.) The strength α associated with the dilaton can be computed, but the range λ at present can only be constrained by experiment; little is known about the mechanism which gives the dilaton mass. The radion results from a very different scenario in which one or more dimensions compactify at TeV energy scales, but here a range λ accessible to our experiment is again predicted.

The special significance of millimetre scales derives from a mass formula of the form $m \approx M^2/M_P$, where M_P is the mass associated with the Planck scale, 10^{19} GeV/ c^2 , and M is a mass of 1–100 TeV/ c^2 , leading to λ in the centimetre to micrometre range. In the strange and gluon moduli case, M is the scale where supersymmetry breaking occurs, whereas in the radion modulus case it corresponds to the size of one or more compact dimensions. This formula also applies to the ADD theory¹⁴ (named for its authors) in which two

compact extra dimensions of millimetre size modify gravity itself. In the ADD theory M is the fundamental length scale where all physics is unified while m is a mass corresponding to the size of the large compact extra dimensions. Many scenarios of compactification, symmetry breaking and mass generation are still viable, so although the possibility of observing new forces at millimetre scales is exciting, such experiments cannot currently falsify string theory.

The planar geometry of our source and detector masses (Fig. 1) is chosen to concentrate as much mass density as possible at the length scale of interest. It is approximately null with respect to the $1/r^2$ newtonian background, a helpful feature in the context of a new force search. A cantilever mode (similar to the motion of a diving board) of the tungsten source mass is driven to a tip amplitude of 19 μm at the resonant frequency of the detector mass. The tungsten detector mass is a double torsional oscillator¹⁵; in the resonant mode of interest the two rectangular sections of the detector counter-rotate about the torsion axis, with most of the amplitude confined to the smaller rectangle under the source mass and shield. Torsional motion of the detector will be driven if a mass coupled force is present between the source and detector. The motions are detected with a capacitive transducer, followed by a preamplifier, filters and a lock-in amplifier. To suppress background forces due to electrostatics and residual gas, a stiff conducting shield is fixed between the test masses. With the source at rest, the gap between source and detector is adjusted to 108 μm , and the entire apparatus is placed in a vacuum enclosure and maintained at pressures below 2×10^{-7} torr.

Figure 2 shows histograms of the raw data, collected with the source mass drive tuned both on and off the detector resonant frequency. Each plot contains data from one channel of the lock-in amplifier, corresponding to one of two orthogonal phases of the

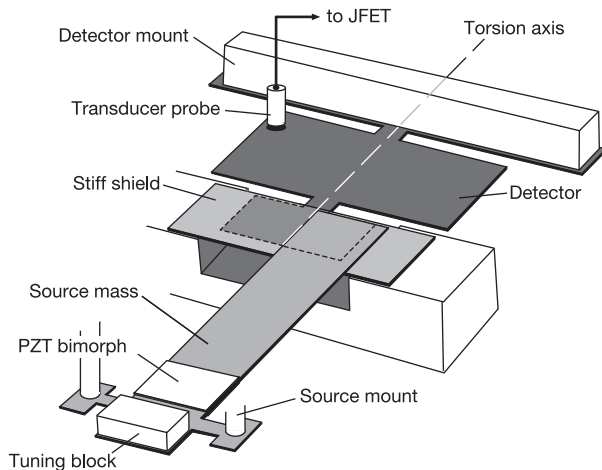


Figure 1 Major components of the apparatus. The smaller rectangle of the tungsten detector (under the shield) is 11.455 mm wide, 5.080 mm long and 195 μm thick. The detector is annealed at 1,300 $^{\circ}\text{C}$ to increase its mechanical Q to 25,000. In operation the 1,173.085-Hz resonant frequency of the detector is stabilized by actively controlling the detector temperature to 305 K. The tungsten source mass is 35 mm long, 7 mm wide and 305 μm thick. The source mass resonant frequency is tuned to the detector and driven by the PZT (lead zirconate titanate) piezoelectric bimorph. The shield is a 60- μm -thick sapphire plate coated with 100 nm of gold on both sides. The test masses and the shield are supported by three separate five-stage passive vibration isolation stacks²³, each providing approximately 200-dB attenuation at 1 kHz. Mechanical probes are used to directly measure the relative orientation and position of each component, and to measure the source mass amplitude. Detector motions are sensed by a cylindrical capacitive probe supported 100 μm above a rear corner of the large rectangle of the detector mass. The probe is biased at 200 V through a 100 G Ω resistor, and connected to an SK 152 junction field-effect transistor (JFET) through a blocking capacitor. The JFET noise temperature of 100 mK is more than sufficient to detect 305 K thermal motions of the detector mass. The JFET preamplifier is followed by a second preamplifier, filters, and finally a two-phase lock-in amplifier. The total voltage gain from the capacitive probe to the lock-in input is approximately 1600. A crystal-controlled oscillator provides a reference signal for the lock-in amplifier and drives the source mass PZT through a 1:10 step-up transformer.

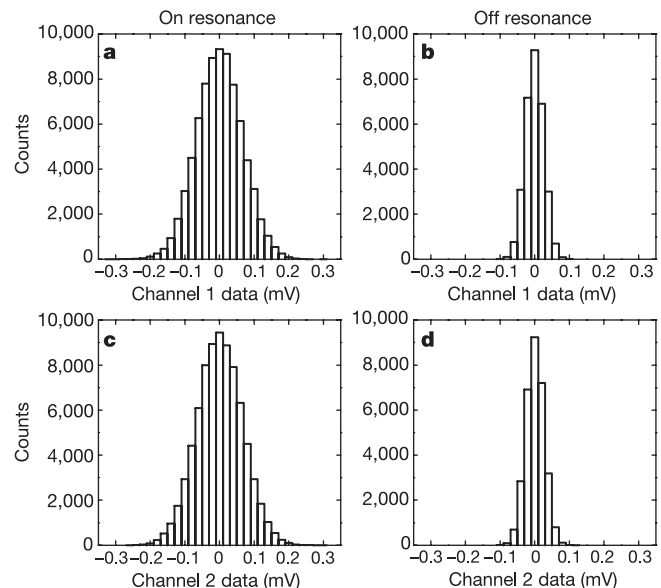


Figure 2 Distributions of data samples. Data were recorded at 1 Hz with a lock-in bandwidth chosen to include the noise power of the detector thermal oscillations, which was used for calibration. Each data cycle began with five 120-sample diagnostic runs with a direct-current bias of 5–10 V applied to the shield to induce a large test force, transmitted from source to detector via deflections of the shield. The biased runs were recorded at five drive frequencies separated by 15 mHz to cover the detector resonance. The shield was then grounded and the cycle continued with 720 samples with the drive tuned on-resonance (a, c) and 288 samples with the drive tuned 2 Hz below the detector resonance (b, d). The off-resonance run provided a continuous zero check. A total of 108 such cycles were acquired over five days yielding 77,760 on-resonance samples. The biased diagnostic data show that the source mass amplitude, the detector Q and resonant frequency, and the electronic gain were all stable throughout the data set.

detector motion at the drive frequency. The widths of the off-resonance distributions are due to preamplifier noise, whereas the on-resonance distributions are due to the sum of preamplifier noise and detector thermal motions. The on- and off-resonance means shown in Fig. 3 agree within their standard deviations, indicating the absence of a significant resonant force signal.

Additional cycles of data were acquired with the source mass on the opposite side of the detector, with a larger, 1-mm gap, and with reduced overlap between the source and detector. No resonant signal was observed in any of these sessions, making it unlikely that the observed null result is due to a fortuitous cancellation of surface potential, magnetic, and/or acoustic effects. Several on-resonance runs were acquired with different transducer probe bias voltage settings. The observed linear dependence on bias voltage of the root-mean-square (r.m.s.) fluctuations in these data is consistent with detector motion due only to thermal noise and rules out additional motion from transducer back-action noise. This check is important because the magnitude of the detector thermal motion is used for calibration.

Data from diagnostic runs with a bias voltage applied to the shield can be used to estimate the minimum size of the residual potential difference between the shield and the (grounded) test masses needed to produce a resonant signal. We find that at least 1.5 V would be needed to generate an effect above detector thermal noise, about an order of magnitude larger than the residual potential difference actually measured between the shield and test masses. The most important magnetic background effect involves eddy currents generated when the source mass moves in an external magnetic field. Fields produced by the source currents create eddy currents in the detector, which then interact with the external field. Studies of this effect with large applied magnetic fields show that the ambient field actually present cannot generate a signal greater than one-fifth of the thermal-noise-limited sensitivity.

The instrument can be calibrated in several ways, but the most accurate method is to use the r.m.s. thermal motion of the detector, which dominates the on-resonance distributions in Fig. 2. According to equipartition, the average kinetic energy in each normal mode of the detector is equal to $\frac{1}{2}kT$, where k is Boltzmann's

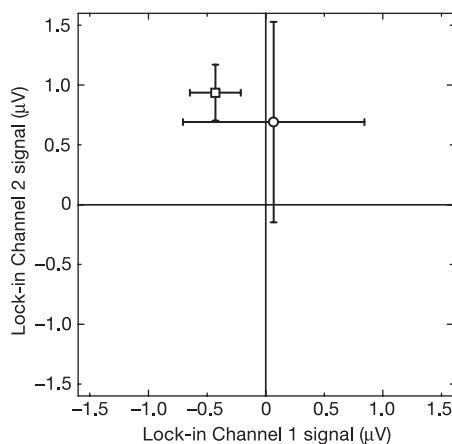


Figure 3 Means of the off- and on-resonance data samples. The circular point with the larger standard deviations is the on-resonance mean. Correlations between nearby samples have been accounted for in computing the standard deviations shown by error bars. The small offset from the origin is due to leakage of the reference signal internal to the lock-in. Measurements with the shield removed and a bias voltage between the source and detector show that the phase for an attractive force is 189° . The on-resonance mean minus the off-resonance mean at 189° is $-0.44 \pm 0.82 \mu\text{V}$. Based on calibration via the equipartition theorem this corresponds to a lumped force amplitude at the edge of the detector (smaller rectangle in Fig. 1) of $-1.2 \pm 2.2 \text{ fN}$, where the negative sign indicates repulsion.

constant and T is the temperature. The normal mode amplitude corresponding to the thermal energy can be calculated, and by comparing this with the observed voltage fluctuations a calibration can be established relating mode amplitude to voltage. A further calculation must be done to find the mode amplitude resulting from any hypothesized force.

For given values of α and λ the driving force due to equation (1) is computed at 30 values of the source mass phase using Monte Carlo integration, and then the Fourier amplitude of the driving force at the resonant frequency is computed. Using the observed statistics of the data we construct a likelihood function for α for each value of λ and compute 95% confidence-level upper limits on α , assuming a uniform prior probability density function (PDF) for α . This analysis is complicated by the presence of uncertainties in the geometrical and mechanical parameters needed to compute the driven displacement. To include these effects we actually construct a likelihood function of α and the uncertain parameters, and then integrate out the uncertain parameters using prior PDFs based on their experimental uncertainties. The most important of these parameters is the $108\text{-}\mu\text{m}$ equilibrium gap between source and detector which has a $6\text{-}\mu\text{m}$ uncertainty. Further details of our analysis methods are given elsewhere¹⁶.

Our results are shown in Fig. 4, together with other experimental limits and the theoretical predictions. Our limit is the strongest available between 10 and $100 \mu\text{m}$. When the gluon and strange moduli forces were first proposed, the areas of their allowed regions in the (α, λ) parameter space were 4.4 and 5.2 square decades respectively. By now they are nearly excluded with only 0.75 and 1.4 square decades still available. Our limit on the dilaton range for $\alpha = 2,000$ is $\lambda < 23 \mu\text{m}$ (corresponding to $m > 8.6 \text{ meV}$), a factor of two better than the previous limit. For the radion modulus we set an upper limit on λ of $88 \mu\text{m}$, close to the value of $40 \mu\text{m}$ estimated

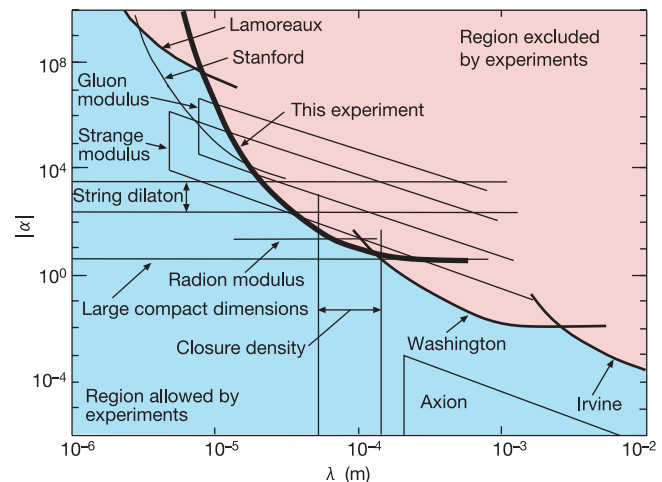


Figure 4 Current limits on new gravitational strength forces between $1 \mu\text{m}$ and 1 cm . Our result is a 95% confidence-level upper limit on the Yukawa strength α as a function of range λ (solid bold curve). It is shown together with limits from previous experiments (Lamoreaux²⁴, Washington¹², Irvine²⁵) and theoretical predictions newly constrained (gluon modulus⁴, strange modulus⁴, string dilaton², radion modulus⁷). An unpublished limit from the Stanford experiment²⁶ is also shown; it is derived in the presence of a background force. The dilaton strength is somewhat model-dependent and there is a range of values reported in the literature². We have chosen the region $200 < \alpha < 3,000$, which includes most values. Also shown are predictions from the ADD theory with two large compact extra dimensions^{14,27,28}, axion mediated forces^{20,21}, and the λ region corresponding to a cosmological energy density between 1.0 and 0.1 times the closure density¹⁷. For the moduli, dilaton, and ADD theories, the upper bounds on λ of the regions shown are set at the experimental limits that were known at the time the theories were proposed.

in the theory (for one extra dimension). For the ADD theory with two large extra compact dimensions, we do not quite reach the limit on the size of these dimensions already set in refs 11 and 12.

Besides forces from extra dimensions, two other ideas have suggested new weak forces at submillimetre scales. The cosmological energy density needed to close the universe, if converted to a length by taking its inverse fourth root (in natural units where $\hbar = c = 1$), corresponds to about 100 μm . This fact has led to repeated attempts^{17–19} to address difficulties connected with the very small observed size of Einstein's cosmological constant by introducing new forces with a range near 100 μm . Our result is the best upper bound on α in this region, but we have not quite reached gravitational sensitivity. Finally, the oldest of these predictions, still out of reach, is the very feeble axion-mediated force^{20,21}. The axion is a field intended to explain why the violation of charge-parity symmetry is so small in quantum chromodynamics, the theory of the strong nuclear force.

Experiments of the sort reported here constrain string-inspired scenarios by setting very restrictive limits on predicted submillimetre forces. Of course, the actual observation of any new force would be a major advance. Because several theoretical scenarios point especially to these length scales, it is an important goal for the future to reach gravitational strength at even shorter distances, perhaps down to 10 μm . Experiments attempting to reach such distances will confront rapidly increasing background forces, especially electrostatic forces arising from the spatially non-uniform surface potentials of metals²². Electric fields due to surface potentials can be shielded with good conductors, but because of the finite stiffness of any shield they still cause background forces to be transmitted between test masses. Stretched membranes (as used by the Washington group) are more effective than stiff plates at the shortest distances, but it remains to be seen down to what distance the background forces can be effectively suppressed. □

Received 21 October 2002; accepted 13 January 2003; doi:10.1038/nature01432.

- Greene, B. *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory* (Norton, New York, 1999).
- Kaplan, D. B. & Wise, M. B. Couplings of a light dilaton and violations of the equivalence principle. *J. High Energy Phys.* **8**, 37 (2000).
- Taylor, T. R. & Veneziano, G. Dilaton couplings at large distances. *Phys. Lett. B* **213**, 450–454 (1988).
- Dimopoulos, S. & Giudice, G. Macroscopic forces from supersymmetry. *Phys. Lett. B* **379**, 105–114 (1996).
- Antoniadis, I. A possible new dimension at a few TeV. *Phys. Lett. B* **246**, 377–384 (1990).
- Antoniadis, I., Dimopoulos, S. & Dvali, G. Millimeter-range forces in superstring theories with weak-scale compactification. *Nucl. Phys. B* **516**, 70–82 (1998).
- Chacko, Z. & Perazini, E. Extra dimensions at the weak scale and deviations from Newtonian gravity. Preprint hep-ph/0210254 available at (arXiv.org) (2002).
- Fischbach, E. & Talmadge, C. *The Search for Non-Newtonian Gravity* (Springer, New York, 1999).
- Bordag, M., Mohideen, U. & Mostepanenko, V. M. New Developments in the Casimir effect. *Phys. Rep.* **353**, 1–205 (2001).
- Long, J. C., Chan, H. W. & Price, J. C. Experimental status of gravitational-strength forces in the sub-centimeter regime. *Nucl. Phys. B* **539**, 23–34 (1999).
- Hoyle, C. D. *et al.* Sub-millimeter tests of the gravitational inverse-square law: A search for “large” extra dimensions. *Phys. Rev. Lett.* **86**, 1418–1421 (2001).
- Adelberger, E. G. Sub-mm tests of the gravitational inverse-square law. Preprint hep-ex/0202008 available at (arXiv.org) (2002).
- Fischbach, E., Krause, D. E., Mostepanenko, V. M. & Novello, M. New constraints on ultrashort-ranged Yukawa interactions from atomic force microscopy. *Phys. Rev. D* **64**, 075010 (2001).
- Arkani-Hamed, N., Dimopoulos, S. & Dvali, G. The hierarchy problem and new dimensions at a millimeter. *Phys. Lett. B* **429**, 263–272 (1998).
- Kleiman, R. N., Kaminsky, G. K., Reppe, J. D., Pindak, R. & Bishop, D. J. Single-crystal silicon high-Q torsional oscillators. *Rev. Sci. Instrum.* **56**, 2088–2091 (1985).
- Long, J. C. *et al.* New experimental limits on macroscopic forces below 100 microns. Preprint hep-ph/0210004 available at (arXiv.org) (2002).
- Beane, S. R. On the importance of testing gravity at distances less than 1 cm. *Gen. Rel. Grav.* **29**, 945–951 (1997).
- Sundrum, R. Towards an effective particle-string resolution of the cosmological constant problem. *J. High Energy Phys.* **7**, 1 (1999).
- Schmidhuber, C. Old puzzles. Preprint hep-th/0207203 available at (arXiv.org) (2002).
- Moody, J. E. & Wilczek, F. New macroscopic forces? *Phys. Rev. D* **30**, 130–138 (1984).
- Rosenberg, L. J. & van Bibber, K. A. Searches for invisible axions. *Phys. Rep.* **325**, 1–39 (2000).
- Price, J. C. in *Proc. Int. Symp. on Experimental Gravitational Physics* (eds Michelson, P., En-ke, H. & Pizzella, G.) 436–439 (World Scientific, Singapore, 1988).
- Chan, H. W., Long, J. C. & Price, J. C. Taber vibration isolator for vacuum and cryogenic applications. *Rev. Sci. Instrum.* **70**, 2742–2750 (1999).

- Lamoreaux, S. K. Demonstration of the Casimir force in the 0.6 to 6 μm range. *Phys. Rev. Lett.* **78**, 5–8 (1997).
- Hoskins, J. K., Newman, R. D., Spero, R. & Shultz, J. Experimental tests of the gravitational inverse-square law for mass separations from 2 to 105 cm. *Phys. Rev. D* **32**, 3084–3095 (1985).
- Chiaverini, J., Smullin, S. J., Geraci, A. A., Weld, D. M. & Kapitulin, A. New experimental constraints on non-Newtonian forces below 100 microns. Preprint hep-ph/0209325 available at (arXiv.org) (2002).
- Floratos, E. G. & Leontaris, G. K. Low scale unification, Newton's law and extra dimensions. *Phys. Lett. B* **465**, 95–100 (1999).
- Kehagias, A. & Sfetsos, K. Deviations from the $1/r^2$ Newton law due to extra dimensions. *Phys. Lett. B* **472**, 39–44 (2000).

Acknowledgements We thank E. Lagae for work in the laboratory, and C. Briggs, T. Buxkemper, L. Czaia, H. Green, S. Gustafson and H. Rohner of the University of Colorado and JILA instrument shops for technical assistance. We also gratefully acknowledge discussions with S. de Alwis, B. Dobrescu and S. Dimopoulos. This work is supported by grants from the US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C.P. (e-mail: john.price@colorado.edu).

Ultra-high-*Q* toroid microcavity on a chip

D. K. Armani, T. J. Kippenberg, S. M. Spillane & K. J. Vahala

Department of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA

The circulation of light within dielectric volumes enables storage of optical power near specific resonant frequencies and is important in a wide range of fields including cavity quantum electrodynamics^{1,2}, photonics^{3,4}, biosensing^{5,6} and nonlinear optics^{7–9}. Optical trajectories occur near the interface of the volume with its surroundings, making their performance strongly dependent upon interface quality. With a nearly atomic-scale surface finish, surface-tension-induced microcavities such as liquid droplets or spheres^{10–13} are superior to all other dielectric microresonant structures when comparing photon lifetime or, equivalently, cavity *Q* factor. Despite these advantageous properties, the physical characteristics of such systems are not easily controlled during fabrication. It is known that wafer-based processing¹⁴ of resonators can achieve parallel processing and control, as well as integration with other functions. However, such resonators-on-a-chip suffer from *Q* factors that are many orders of magnitude lower than for surface-tension-induced microcavities, making them unsuitable for ultra-high-*Q* experiments. Here we demonstrate a process for producing silica toroid-shaped microresonators-on-a-chip with *Q* factors in excess of 100 million using a combination of lithography, dry etching and a selective reflow process. Such a high *Q* value was previously attainable only by droplets or microspheres and represents an improvement of nearly four orders of magnitude over previous chip-based resonators.

Devices were fabricated upon silicon wafers prepared with a 2- μm layer of silicon dioxide (SiO_2). The fabrication process flow is composed of four steps: photolithography; pattern transfer into the silicon dioxide layer; selective, dry etch of the exposed silicon; and selective reflow of the patterned silica. The process details are as follows (Fig. 1). First, photolithography is performed to create disk-shaped photo-resist pads (160 μm in diameter) on a (100) prime grade silicon substrate with 2 μm of oxide grown using wet thermal oxidation in a horizontal tube furnace. An additional bake follows in order to reflow the photo-resist, smoothing the edges in the

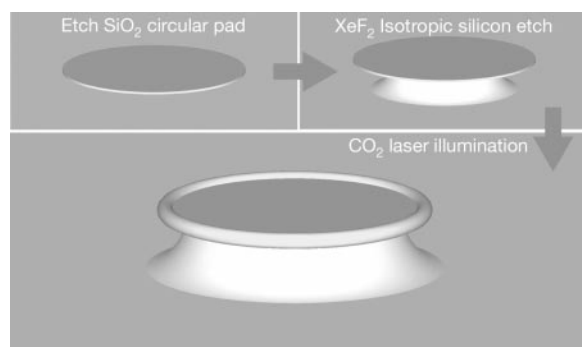


Figure 1 Flow diagram illustrating the process used to fabricate ultra-high- Q planar microcavities.

process. The circular disks of photo-resist act as an etch mask during immersion in buffered HF solution at room temperature. Acetone is then used to remove residual photo-resist and organic contamination. The remaining SiO_2 disks act as etch masks during exposure to XeF_2 gas at 3 torr. XeF_2 was specifically chosen for the purpose of isotropic selective removal of silicon. As a result, the edges of the SiO_2 disks are equally undercut, leaving circular silicon pillars supporting larger SiO_2 disks. As the optical trajectories of interest reside at the periphery of the silica disk, the removal of the higher-index silicon below a portion of the periphery is required to inhibit power leakage into the substrate. We observe Q values in excess of 10^5 for these structures using the spectral characterization approach described below. The above process flow leaves lithographic blemishes, visible in an optical microscope, at the all-important disk periphery. Therefore we used additional processing to achieve the surface finish characteristic of surface-tension-induced microcavity structures (root-mean-square (r.m.s) roughness of several nanometres or less¹³).

Next, a processing step is introduced to selectively heat and reflow the undercut SiO_2 disks without affecting the underlying silicon support pillar. An undercut SiO_2 disk is surface-normal-irradiated using a CO_2 laser (10.6- μm wavelength), similar to techniques proposed for integrated circuit planarization¹⁵. The beam intensity profile follows an approximately gaussian distribution and is focused to a circular spot approximately 200 μm in diameter. The resulting beam intensity can be varied by electronic control of the laser power, but is typically 100 MW m^{-2} during reflow. Owing to the strong temperature dependence of the silica optical extinction coefficient near 10.6 μm (ref. 16) as well as the thermal isolation of the undercut SiO_2 disk, melting of the disk occurs along the periphery, and significantly not over the silicon pillar. In addition to having a far weaker optical absorption at 10.6 μm , silicon is 100 times more thermally conductive than silica^{16,17}. The silicon pillar therefore remains significantly cooler and physically unaffected throughout the silica reflow process, serving as a heat sink to the selectively absorbed optical power in the silica layer. As the disk diameter shrinks, the effective cross-section available to absorb laser power decreases and shrinkage is observed to terminate when a toroid-like silica structure has formed. Beyond this point, continued laser treatment at the same intensity results in no observable change of the structure. The process is therefore self-quenching with the final diameter of the molten disk rim controlled by lithography and chemical etch steps. It should be noted that it is possible to interrupt the reflow prior to quenching thereby producing a toroid with a diameter intermediate to the initial disk diameter and terminal diameter.

Thus, in this process step, surface tension both smoothes the surface and collapses the silica disk to a toroid shape with limiting dimensions defined by the support pillar and the initial thickness of the silica layer. Micrographs showing disks both before and after the

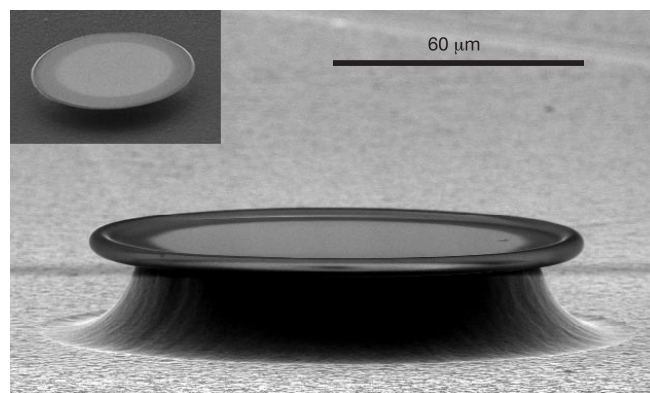


Figure 2 Scanning electron micrograph of a silica microdisk after selective reflow treatment with a CO_2 laser. The inset shows the microdisk prior to laser treatment. This toroidal microresonator had an intrinsic cavity Q of 1.00×10^8 .

laser-activated selective reflow process are shown in Fig. 2. In these micrographs, the overall disk diameter was reduced to 120 μm as silica was consumed to form a 7- μm -thick toroid-shaped perimeter.

The mode structure and quality factor of the toroidal cavities were characterized in the optical telecommunication band (1,500 nm wavelength band). Tapered optical fibres connected to a single-mode, tunable, external-cavity laser were used to efficiently excite 'whispering gallery' modes of the resonators. Tapered waveguides were positioned on a 20-nm-resolution stage and could be moved freely over the silicon sample to couple individually to each of the toroid-shaped microresonators. Dual microscopes were used to simultaneously image disk resonators and fibre tapers from the side and the top. Proper alignment required the taper axis to reside in the equatorial plane of the toroidal cavity with minimal tilt angle. Briefly, taper fabrication¹⁸ proceeds by stretching a standard optical fibre (SMF-28), which has been heated using a hydrogen flame. By observing the adiabatic condition, the tapered fibres exhibit low fibre-to-fibre insertion loss (typically $<10\%$). Taper waist diameters are typically several micrometres, adjusted to properly phase-match to the resonator. Critical coupling¹⁹ (the resonant transfer of all optical waveguide power into the resonator) was achieved by appropriately adjusting the taper-resonator gap. Non-resonant loss was observed to be low ($<5\%$).

Figure 3 shows the transmission spectra through a taper in close proximity (of the order of hundreds of nanometres) to a 94- μm -diameter toroidal microresonator. The observed free spectral range corresponds to the equatorial mode number (l -index). Inspection of the data shows that the resonator supports very few radial and azimuthal (m -index or transverse) modes. This is in contrast to spheres, which support $(2l + 1)$ azimuthal modes. The quality factor, or Q , of the resonators was measured in two ways. First, the full-width at half-maximum (FWHM) of the lorentzian-shaped resonance in the undercoupled regime was directly measured by scanning a single-mode laser (short-term linewidth about 300 kHz) through a resonance. Low input power levels (typically less than 5 μW) were used to avoid thermally induced distortion of the line shape caused by resonant-field build-up within the cavity. Repeated measurements on samples fabricated with various radii (80–120 μm) and tori thickness (5–10 μm) yielded Q values in excess of 100 million (10^8). This is a record value for a planar device, and constitutes an improvement of nearly four orders of magnitude over all previous planar microresonators fabricated by wafer-scale processing (the highest values^{20,21} reported thus far are less than 2×10^4).

As an independent and more precise measurement of quality factor, the photon lifetime was directly measured by cavity ring-down. This was done by repeatedly scanning the laser into resonance with a mode that was critically coupled to the taper. As the laser

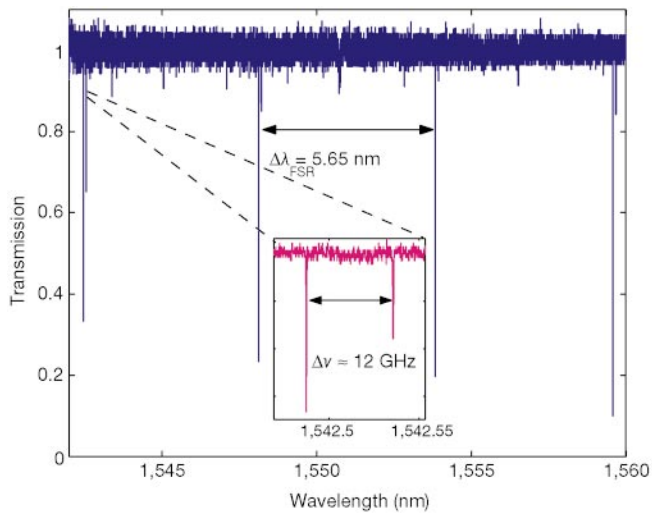


Figure 3 Transmission spectra of a toroidal resonator. The free spectral range (defined as the wavelength spacing between modes with successive angular mode number) is 5.65 nm, which corresponds to a torus approximately 94 μm in diameter. The inset shows what we believe to be the two lowest-order radial modes (based on modelling of a microdisk resonator). Additional subsidiary peaks are attributed to other radial or azimuthal modes.

scanned into resonance, power transfer increased until maximal 'charging' of the resonator mode was attained. At this moment, the laser input was gated 'off' by use of a high-speed, external modulator and cavity ringdown is observed as the resonant power discharges. Because the resonator is by necessity loaded during this measurement, the observed ringdown time yields the cavity lifetime at the critical point, τ_{crit} , and the loaded Q -factor (not the intrinsic quality factor, Q_0). Data from a typical ringdown measurement are shown in Fig. 4. At time $t = 0$ in the figure, a signal is applied to 'gate' the laser off. When the laser input is fully off, the detected power is due entirely to the cavity discharge field. The solid line represents an exponential fit as expected for decay of a single cavity mode. The inset shows a logarithmic plot to infer the cavity lifetime. The loaded lifetime in this structure was 43 ns. As a further check on this time constant, after gating of the pump laser the waveguide power has dropped to 80% of its predicted maximal value based on extrapolation of data to $t = 0$. This value is in agreement with the gating delay of the ringdown setup ($\Delta t \approx 8$ ns). In particular, using the observed mode-lifetime of $\tau_{\text{crit}} = 43$ ns yields $\exp(-\Delta t/\tau_{\text{crit}}) = 0.83$.

As noted before, to infer the intrinsic cavity Q it is necessary to correct for loading by the taper waveguide. In addition, it is necessary to take into account excitation of the counter-propagating mode due to scattering centres intrinsic to the resonator (described by a dimensionless intermode coupling parameter Γ). The techniques used to measure this parameter in ultra-high- Q taper-coupled resonators are described elsewhere²². For the mode of Fig. 4 the intermode coupling was measured to be approximately 1, giving rise to a weak counter-propagating wave excitation (17% of the cavity build-up field is stored in the counter-propagating mode at critical coupling). In the presence of intermode coupling the relationship between the critically coupled lifetime and the intrinsic (unloaded) cavity quality factor, is given by

$$Q_0 \equiv \omega\tau_0 = \omega\tau_{\text{crit}}(1 + \sqrt{1 + \Gamma^2})$$

This yields an intrinsic cavity Q of 1.25×10^8 inferred from cavity ringdown. This value agrees with the measurements of the frequency line shape described above.

Thus we have fabricated ultra-high- Q planar cavities on a chip for the first time (to our knowledge). Toroid-shaped microcavities were

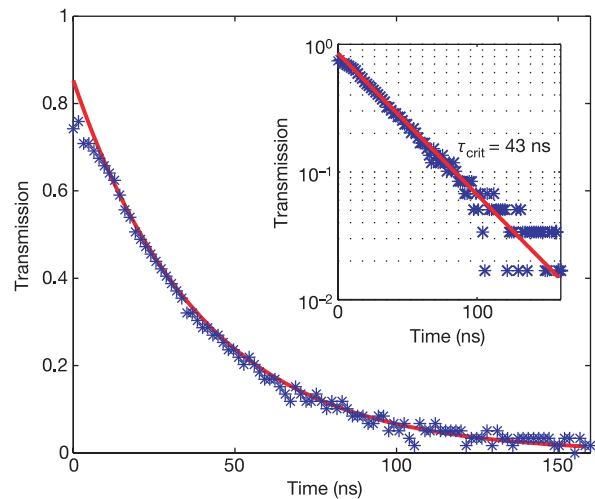


Figure 4 Ringdown measurement of a 90- μm -diameter toroid microcavity at the critical-coupling point. The measured lifetime of $\tau_{\text{crit}} = 43$ ns corresponds to an intrinsic quality factor of $Q = 1.25 \times 10^8$.

formed using a combination of lithography, dry etching and a selective reflow process. Self-limited collapse of a molten silica disk enables the dimensional control typical of wafer-scale processing while providing the surface finish (and hence cavity Q) typical of a spherical resonator. Q values obtained by this process are typically four orders of magnitude higher than previous wafer-based resonators. In some applications mode volume is also an important factor, and certain chip-based micro-resonators^{20,21} feature smaller mode volumes than the present structures. As a gauge to the lower bound on size, radiation leakage becomes a significant factor in determining Q for diameters less than 20 μm in silica microspheres. It should therefore be possible to reduce toroid diameter (and hence mode volume) substantially from the present 100 μm while maintaining ultra-high- Q values. The planar nature of the toroid microcavity and the large transparency window of silica suggest that these devices will find a wide range of applications in photonics as well as in fundamental studies. As an indication of the possibilities for these structures, in the course of this work nonlinear optical effects have been observed with characteristics comparable to recent studies on spherical ultra-high- Q cavities⁹. As standard processing techniques are used, the addition of optical functionality by techniques such as implantation or coating is possible. Likewise, electrical functionality can be introduced to integrate control functions with the ultra-high- Q microcavities. More generally, this work provides a new functional element that is synergistic with recent demonstrations of basic experimental physics on a chip. For example, by combining the present results with techniques recently demonstrated to integrate atomic traps on a chip²³ it would be possible to achieve chip-scale integration of cavity quantum electrodynamics experiments and related devices. Finally, there is great interest in improving the sensitivity of biological and chemical sensors. Proposals based upon optical resonators will benefit from the ability to attain ultra-high Q on a chip. □

Received 10 October; accepted 16 December 2002; doi:10.1038/nature01371.

1. Lefevre-Seguin, V. & Haroche, S. Towards cavity-QED experiments with silica microspheres. *Mater. Sci. Eng. B* **48**, 53–58 (1997).
2. Vernooij, D. W., Furusawa, A., Georgiades, N. P., Ilchenko, V. S. & Kimble, H. J. Cavity QED with high- Q whispering gallery modes. *Phys. Rev. A* **57**, R2293–R2296 (1998).
3. McCall, S. L., Levi, A. F. J., Slusher, R. E., Pearton, S. J. & Logan, R. A. Whispering-gallery mode microdisk lasers. *Appl. Phys. Lett.* **60**, 289–291 (1992).
4. Sandoghdar, V. et al. Very low threshold whispering-gallery-mode microsphere laser. *Phys. Rev. A* **54**, R1777–R1780 (1996).
5. Vollmer, F. et al. Protein detection by optical shift of a resonant microcavity. *Appl. Phys. Lett.* **80**, 4057–4059 (2002).

6. Serpenguzel, A., Arnold, S. & Griffl, G. Excitation of resonances of microspheres on an optical fiber. *Opt. Lett.* **20**, 654–656 (1995).
7. Chang, R. K. & Campillo, A. J. (eds) *Optical Processes in Microcavities* (World Scientific, Singapore, 1996).
8. Treussart, F. *et al.* Evidence for intrinsic Kerr bistability of high-Q microsphere resonators in superfluid helium. *Eur. Phys. J. D* **1**, 235–238 (1998).
9. Spillane, S. M., Kippenberg, T. J. & Vahala, K. J. Ultralow-threshold Raman laser using a spherical dielectric microcavity. *Nature* **415**, 621–623 (2002).
10. Campillo, A. J., Eversole, J. D. & Lin, H.-B. Cavity quantum electrodynamic enhancement of stimulated emission in microdroplets. *Phys. Rev. Lett.* **67**, 437–440 (1991).
11. Braginsky, V. B., Gorodetsky, M. L. & Ilchenko, V. S. Quality-factor and nonlinear properties of optical whispering-gallery modes. *Phys. Lett. A* **137**, 393–397 (1989).
12. Collot, L., Lefevre-Seguin, V., Brune, M., Raimond, J. M. & Haroche, S. Very high-Q whispering-gallery mode resonances observed on fused-silica microspheres. *Europhys. Lett.* **23**, 327–334 (1993).
13. Vernooy, D. W., Ilchenko, V. S., Mabuchi, H., Steed, E. W. & Kimble, H. J. High-Q measurements of fused-silica microspheres in the near infrared. *Opt. Lett.* **23**, 247–249 (1998).
14. Himeno, A., Kato, K. & Miya, T. Silica-based planar lightwave circuits. *IEEE J. Sel. Top. Quant.* **4**, 913–924 (1998).
15. Delfino, M. & Reifsteck, T. A. Laser activated flow of phosphosilicate glass in integrated circuit devices. *Elect. Device Lett.* **3**, 116–118 (1982).
16. McLachlan, A. D. & Meyer, F. P. Temperature dependence of the extinction coefficient of fused silica for CO₂ laser wavelengths. *Appl. Opt.* **26**, 1728–1731 (1987).
17. Sheik-bahae, M. & Kwok, H. S. Controlled CO₂ laser melting of silicon. *J. Appl. Phys.* **63**, 518–524 (1988).
18. Knight, J. C., Cheung, G., Jacques, F. & Birks, T. A. Phase-matched excitation of whispering-gallery-mode resonances by a fiber taper. *Opt. Lett.* **22**, 1129–1131 (1997).
19. Cai, M., Painter, O. & Vahala, K. J. Observation of critical coupling in a fiber taper to a silica-microsphere whispering-gallery mode system. *Phys. Rev. Lett.* **85**, 74–77 (2000).
20. Michler, P. *et al.* Laser emission from quantum dots in microdisk structures. *Appl. Phys. Lett.* **77**, 184–186 (2001).
21. Gayral, B. *et al.* High-Q wet-etched GaAs microdisks containing InAs quantum boxes. *Appl. Phys. Lett.* **75**, 1908–1910 (1999).
22. Kippenberg, T. J., Spillane, S. M. & Vahala, K. J. Modal coupling in traveling-wave resonators. *Opt. Lett.* **27**, 1669–1671 (2002).
23. Hansel, W., Hommelhoff, P., Hansch, T. W. & Reichel, J. Bose–Einstein condensation on a microelectronic chip. *Nature* **413**, 498–501 (2001).

Acknowledgements This work was supported by DARPA and the Caltech Lee Center.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.J.V. (e-mail: vahala@caltech.edu).

Logarithmic rate dependence of force networks in sheared granular materials

R. R. Hartley & R. P. Behringer

Department of Physics & Center for Nonlinear and Complex Systems, Duke University, Durham, North Carolina 27708-0305, USA

Many models of slow, dense granular flows^{1–5} assume that the internal stresses are independent of the shearing rate. In contrast, logarithmic rate dependence is found in solid-on-solid friction^{6–8}, geological settings^{9–11} and elsewhere^{12–15}. Here we investigate the rate dependence of stress in a slowly sheared two-dimensional system of photoelastic disks, in which we are able to determine forces on the granular scale. We find that the mean (time-averaged) stress displays a logarithmic dependence on the shear rate for plastic (irreversible) deformations. However, there is no perceivable dependence on the driving rate for elastic (reversible) deformations, such as those that occur under moderate repetitive compression. Increasing the shearing rate leads to an increase in the strength of the force network and stress fluctuations. Qualitatively, this behaviour resembles the changes associated with an increase in density. Increases in the shearing rate also lead to qualitative changes in the distributions of stress build-up and relaxation events. If shearing is suddenly stopped, stress

relaxations occur with a logarithmic functional form over long timescales. This slow collective relaxation of the stress network provides a mechanism for rate-dependent strengthening.

Slow granular flows are typically described in the context of Mohr–Coulomb friction models² that resemble those used for describing friction between two solid bodies¹⁶. In the well-known solid friction scheme^{17,18}, an object on a frictional surface will resist a force and remain at rest, provided that the magnitude of the tangential force is less than the product of a static friction coefficient and the normal force. This picture was translated into the granular context (for dense granular systems characterized by networks of force chains¹⁹) by Coulomb²⁰ and more recent authors^{1,2}. In this context the normal and tangential forces are replaced by corresponding normal and shear stresses, and the surface of interaction is replaced by a plane within the material. For large enough tangential force relative to the normal force sliding friction (failure in a granular material) occurs. In these pictures, sliding friction (deformation following failure) is independent of the speed of sliding (the shear rate).

In reality, experiments in diverse contexts have shown that solid friction exhibits a logarithmic dependence on rate and that static frictional contacts strengthen logarithmically with age. These experiments span a vast range of lengths, and include studies at the atomic¹⁵, laboratory^{7,8,12–14} and geological scale^{9–11}. Recent experiments⁶ using granular materials sliding against the interior wall of a piston showed a clear rate dependence that was associated with ageing effects of individual solid-friction contacts and with the force network. Experiments^{7,8} in which a solid rough surface was pushed across a granular bed showed a slow strengthening (or ageing) with time of the (quasi-static) force, and creep before slip events. Other experiments¹¹ have found strengthening due to ageing for granular samples under large compressions. In the present experiments, which consider the grains and the force network, we show that there is a logarithmic rate-dependence in slowly sheared granular materials that is associated with irreversible rearrangements of the force network within the material itself. This is manifested as a strengthening with rate.

The experiments described here were carried out with a two-dimensional realization of a granular system. The particles were made of a photoelastic material, and were either relatively thin disks or flat particles with a pentagonal cross-section. By using a photoelastic material, we could determine forces at the grain scale. More detailed descriptions of the experimental apparatus and methods²¹

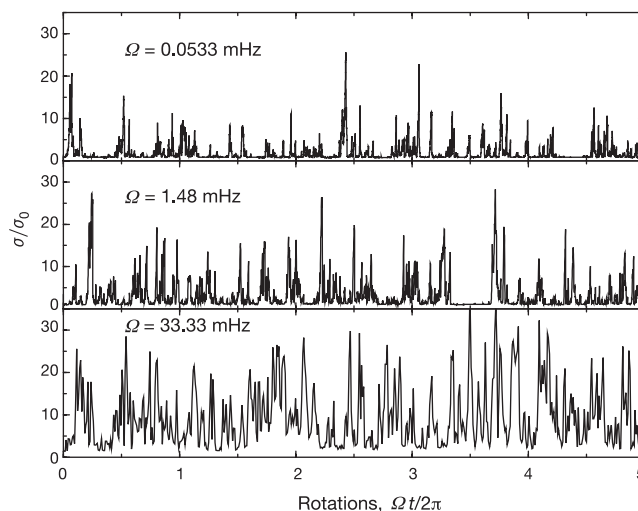


Figure 1 Time series for stress (normalized by $\sigma_0 \approx 4.11 \text{ N m}^{-1}$) for a range of shearing rates Ω that span the experimental range. These data are for pentagonal particles, but data for disks are qualitatively similar. Shear wheel radius 13.8 cm, $\gamma = \gamma_c + 0.0152$.

have been given elsewhere for related experiments^{19,22,23}, and here we provide only essential information for two qualitatively different experiments. In the first of these two experiments, the photoelastic particles were sheared in an annular geometry (Fig. 2 inset) that was bounded on the inside by a rough wheel, the source of shearing, and on the outside by an equally rough static ring. The particles sat on a horizontal, flat powder-lubricated Perspex sheet, and were contained from above by a similar sheet. This assembly was placed in a polariscope consisting of a pair of right and left circular polarizers, a light source and a video camera. Photoelastic images (effective shutter speed 10^{-2} s) were obtained at rates up to 15 Hz, captured by a frame-grabber, and processed in real time to determine the forces on the particles within the field of view of the camera (~ 200 particles). We computed the force on each particle in each frame and summed over all particles in a frame. This integrated force was then stored for each image to produce a time series over very long sampling times, between 2^{14} – 2^{17} points depending on the rate of shearing, Ω , and for 10 to approximately 200 rotations of the shearing wheel, depending on Ω . To a reasonable approximation ($\sim 10\%$), this quantity is proportional to the pressure, and for simplicity, we will refer to it as stress.

Important control parameters are Ω ($=2\pi/T$, where T is the period for one rotation of the shearing wheel) and the density of the system, which we give in terms of the area fraction, γ , occupied by the particles. As shown previously^{19,22}, there is a critical value when $\gamma = \gamma_c$, below which the force network vanishes, and we typically reference γ to this value. (γ_c depends on properties such as the particle shape and possibly the geometry of the container.) It is also useful to give the rotation rate, $f = 1/T = \Omega/2\pi$. In these experiments they were $0.03 \text{ MHz} \leq f \leq 60 \text{ MHz}$.

In Fig. 1 we show several examples of stress time series for values of Ω ranging over about three orders of magnitude. The three images of this figure give an indication of the effect of the shearing rate; the time series for the slowest rate is clearly of lower amplitude and more intermittent than that for the highest rate.

The rate dependence is seen clearly in the mean (time-averaged)

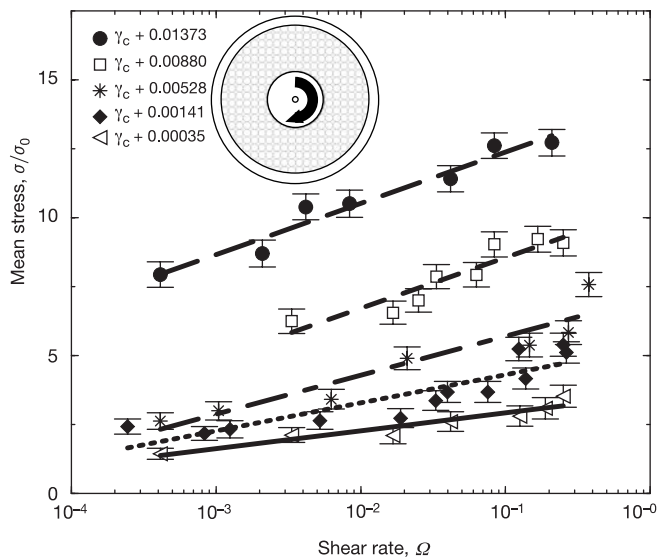


Figure 2 Mean stress versus shearing rate, Ω , for a range of densities for pentagonal particles. Note that these densities are relatively close to the critical value (γ_c) at which the force network vanishes. The linear-logarithmic scales emphasize the fact that the data are consistent with a logarithmic variation of the stress with Ω . Here, $\sigma_0 \approx 4.11 \text{ N m}^{-1}$. Inset, sketch of Couette shear apparatus. The particles used in the experiment had typical linear dimensions between 0.7 cm and 0.9 cm, a Young's modulus of 4 MPa and a density 1.15 g cm^{-3} . For either shear or compression experiments, the humidity varied by no more than $\pm 5\%$.

stress versus Ω in Fig. 2. These data are consistent with a logarithmic rate dependence for all values of γ . Similar effects have been seen in previous experiments^{6,8,11,12} but with a significant difference: in the present experiments, there is a logarithmic growth of the mean stress (that is, the force network) with rate, while other experiments show a logarithmic strengthening of the friction coefficient with age and/or humidity. If we fit these data to the form $\bar{\sigma} = A \log[\Omega/\Omega_0]$, then the amplitude, A , is sensitive to the density, but the characteristic frequency, Ω_0 , depends only weakly on γ , and corresponds to a typical time of $2\pi/\Omega_0 \approx 30 \text{ d}$. However, there is no compelling reason to believe that this relation holds in the limit $\Omega \rightarrow 0$. A more realistic description may be $\bar{\sigma} = A \log[(\Omega + \Omega_1)/\Omega_0]$, where $\Omega_1 > \Omega_0$.

If a sample is sheared under steady-state conditions and the shearing wheel is abruptly stopped, the force network relaxes over very long timescales. In Fig. 3a, we show the changes that have occurred in a particular case after 10 h. In Fig. 3b we show quantitative data for the relaxation of stress versus time on a semi-log plot. Note that there are two timescales: a quicker initial decay lasting about 20–300 s and a much longer process such that the stress network may still be relaxing after 20 h. This behaviour may provide insight into the creep events observed⁷ in granular friction experiments. Figure 3b shows fits to the stress decrease at long times of the form $\sigma(t)/\sigma(t=0) = B \log(t/t_0)$. Although the relaxation is not completely uniform in time, a logarithmic functional form is not unreasonable for many cases (after an initial rapid relaxation). Note an example of a major relaxation event in one of these runs about 350 s after the cessation of shearing: thereafter, continued slow relaxation occurred. In Fig. 3c we contrast relaxation stress data for samples prepared under uniform compression to the relaxation of initially sheared samples in Fig. 3b. In the compression experiments no relaxation is seen within experimental resolution.

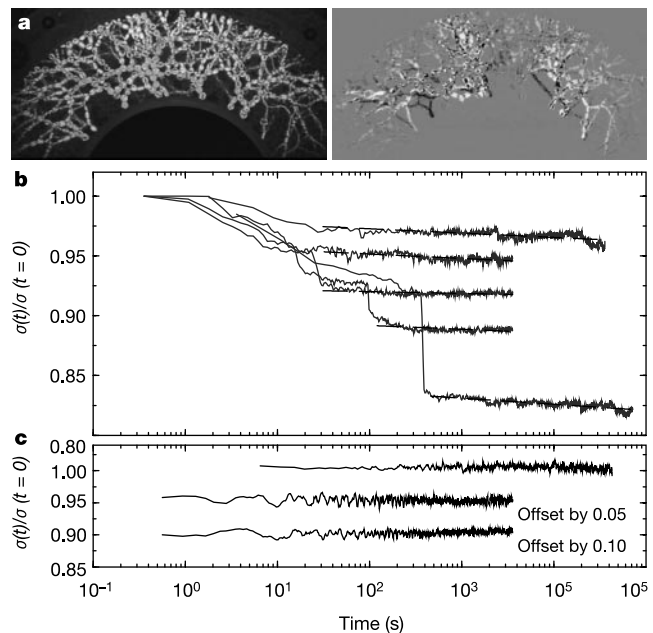


Figure 3 Stress relaxation in sheared and compressed systems. **a**, For the shear experiment, we show the initial force network (left) and the change in force network over a 10-h period (right). In the latter image, decreases and increases in the photoelastic image intensity are shown in white and black, respectively. **b**, Relaxation for particles that were undergoing plastic shear deformation due to shear before $t = 0$. At $t = 0$, the shearing wheel was stopped and maintained in a fixed position. The dashed lines are a fit to $\sigma(t)/\sigma(t=0) = B \log(t/t_0)$. **c**, Similar stress versus time data for a collection of particles that are maintained under uniform compression.

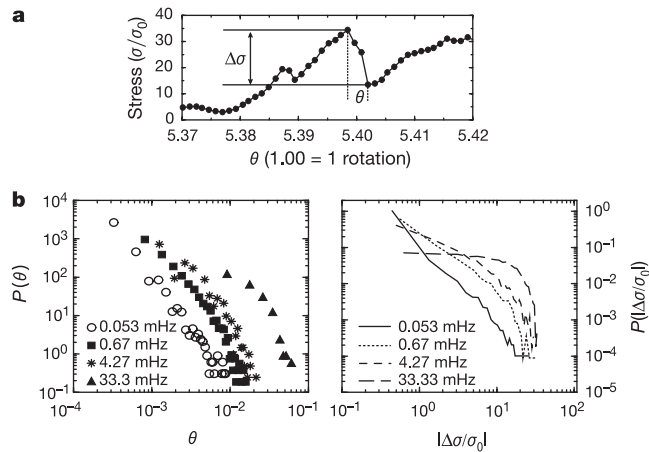


Figure 4 Stress “avalanches” in sheared granular materials. Here, $\sigma_0 \approx 4.11 \text{ N m}^{-1}$; particles were pentagons with $\gamma = \gamma_c + 0.022$. **a**, A short segment of a time series that identifies a single avalanche, $\Delta\sigma$ and θ . **b**, Distributions $P(\theta)$ and $P(|\Delta\sigma/\sigma_0|)$ of stress avalanches for a dense system of pentagons for several shear rates.

We gain additional insight into the effect of rate changes by determining the distribution of build-up events (monotonic increases in stress) and release events (‘avalanches’—monotonic decreases in stress). Here, there are two aspects that are of interest: the size of the build-ups and avalanches, $\Delta\sigma$, and the angle, θ , through which the shearing wheel turns during the course of a build-up or an avalanche. In Fig. 4a, we show an example of an avalanche event, and in Fig. 4b, we show distributions of $\Delta\sigma$ and θ for avalanches. For near-critical γ and for the slowest Ω the avalanche distributions are consistent with the power laws. At higher γ and Ω the distributions for both build-up (not shown) and avalanche events are roughly exponential. We note that increases in both Ω and in γ tend to have the same qualitative effect on the statistics of avalanches and build-ups.

To further explore the origin of this rate dependence, we have carried out a second set of experiments, using the same photoelastic techniques, in which the particles were confined to a box with three fixed sides. The sample was then compressed and released repeatedly on the fourth side by means of an oscillating piston (Fig. 5 inset). The driving amplitude of the piston ($a = 0.44 \text{ cm}$) was chosen so that the typical piston velocity matched the corresponding velocities of the shear wheel in the first experiment. In the first set of experiments there was continuous plastic shear deformation, whereas there was none in the second set—the particles maintained their relative position. Within experimental error, there was no rate dependence of this process (Fig. 5 main figure).

We have demonstrated that sheared granular materials exhibit a logarithmic rate dependence that is tied in an essential way to force network/particle rearrangements. The nature of this rate dependence is further illuminated by the slow relaxation that occurs when shearing is halted. These experiments in particular show that there are slow collective rearrangements of the force network that can occur over large timescales. Although the frictional micro-contacts between particles may be strengthening in time^{8,12,16}, the force network as a whole is relaxing over very long timescales. This slow relaxation of forces is consistent with increases in stress with rate. Force chains generated by shearing cannot completely relax on the timescales over which new chains are formed, an effect that becomes stronger with increasing Ω . It seems likely that this effect is related to the long timescales associated with collective rearrangements in other settings, such as compaction in granular materials²⁴. For compaction, rearrangements become progressively more difficult over time, because each new rearrangement requires the involvement of larger collections of particles. Similarly, in the

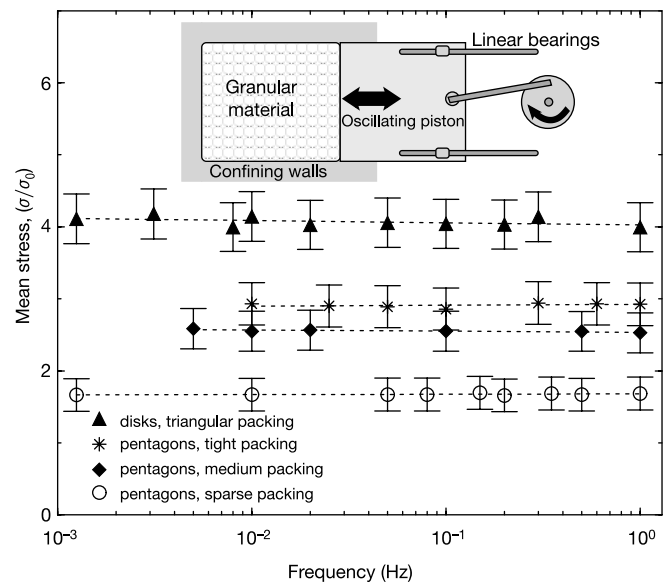


Figure 5 Mean stress versus oscillation rate, ω , for several densities for oscillatory compression. Data for both disks and pentagons are shown. The linear-logarithmic scales emphasize the fact that there is no variation of the stress with ω within experimental uncertainties. $\sigma_0 \approx 3.83 \text{ N m}^{-1}$. Inset, sketch of piston apparatus.

present experiments, relaxation events involving collections of particles become less probable over time owing to geometric constraints on the particles, to the fact that those contacts that are near failure will have slipped, and to the fact that the available elastic energy is gradually reduced with each rearrangement. An interesting question is whether similar properties could be found in other jammed systems²⁵ such as colloids or foams. □

Received 7 October; accepted 30 December 2002; doi:10.1038/nature01394.

- Wood, D. M. *Soil Behaviour and Critical State Soil Mechanics* (Cambridge Univ. Press, Cambridge, UK, 1990).
- Nedderman, R. M. *Statics and Kinematics of Granular Materials* (Cambridge Univ. Press, Cambridge, UK, 1992).
- Jaeger, H. M., Nagel, S. R. & Behringer, R. P. Granular solids, liquids, and gases. *Rev. Mod. Phys.* **68**, 1259–1273 (1996).
- Herrmann, H. J., Hovi, J. P. & Luding, S. (eds) *Physics of Dry Granular Media* (NATO ASI Series, Kluwer, Dordrecht, 1997).
- Behringer, R. P. & Jenkins, J. T. (eds) *Powders and Grains 97* (Balkema, Rotterdam, 1997).
- Overez, G., Kolb, E. & Clément, E. Rheology of a confined granular material. *Phys. Rev. E* **64**, 060302–060305(R) (2001).
- Nasuno, S., Kudrolli, A., Bak, A. & Gollub, J. P. Time-resolved studies of stick-slip friction in sheared granular layers. *Phys. Rev. E* **58**, 2161–2171 (1998).
- Losert, W., Gémard, J.-C., Nasuno, S. & Gollub, J. P. Mechanisms for slow strengthening in granular materials. *Phys. Rev. E* **61**, 4060–4068 (2000).
- Dieterich, J. H. Modeling of rock friction 1: Experimental results and constitutive equations. *J. Geophys. Res.* **84**, 2161–2168 (1979).
- Ruina, A. L. Slip instability and state variable friction laws. *J. Geophys. Res.* **88**, 10359–10370 (1983).
- Karner, S. L. & Marone, C. The effect of shear load on frictional healing in simulated fault gouge. *Geophys. Res. Lett.* **25**, 4561–4564 (1998).
- Berthoud, P., Baumberger, T., G'Sell, C. & Hiver, J. M. Physical analysis of the state- and rate-dependent friction law: Static friction. *Phys. Rev. B* **59**, 14313–14327 (1999).
- Baumberger, T., Berthoud, P. & Caroli, C. Physical analysis of the state- and rate-dependent friction law. II. Dynamic friction. *Phys. Rev. B* **60**, 3928–3939 (1999).
- Carlson, J. M. & Batista, A. A. Constitutive relation for the friction between lubricated surfaces. *Phys. Rev. E* **53**, 4153–4165 (1996).
- Gnecco, E. et al. Velocity dependence of atomic friction. *Phys. Rev. Lett.* **84**, 1172–1175 (2000).
- Heslot, F., Baumberger, T., Perrin, B., Caroli, B. & Caroli, C. Creep, stick-slip, and dry-friction dynamics: Experiments and a heuristic model. *Phys. Rev. E* **49**, 4973–4988 (1994).
- Persson, B. N. J. *Sliding Friction: Physical Principles and Applications* (Springer, Berlin, 1998).
- Meyer, E., Overney, R. M., Dransfeld, K. & Gyalog, T. *Nanoscience, Friction and Rheology on the Nanometer Scale* (World Scientific, Singapore, 1998).
- Howell, D., Veje, C. & Behringer, R. P. Stress fluctuations in a 2D granular Couette experiment: A continuous transition. *Phys. Rev. Lett.* **82**, 5241–5244 (1999).
- Coulomb, A. *Mémoires de Mathématiques et de Physique* (L'Imprimerie Royale, Paris, 1776).
- Heywood, R. B. *Designing by Photoelasticity* (Chapman and Hall, London, 1952).
- Howell, D. W., Behringer, R. P. & Veje, C. T. Fluctuations in granular media. *Chaos* **9**, 559–572 (1999).
- Geng, J. et al. Footprints in sand: The response of a granular material to local perturbations. *Phys. Rev. Lett.* **87**, 035506–035509 (2001).

24. Nowak, E. R., Knight, J. B., Ben-Naim, E., Jaeger, H. M. & Nagel, S. R. Density fluctuations in vibrated granular. *Phys. Rev. E* **57**, 1971–1982 (1998).
25. Liu, A. L. & Nagel, S. R. Nonlinear dynamics: Jamming is just not cool anymore. *Nature* **396**, 21–22 (1998).

Acknowledgements We thank S. Coppersmith for discussion. This work was supported by NSF Division of Materials Research, NSF Division of Mathematical Sciences and by NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.P.B. (e-mail: bob@phy.duke.edu).

Early history of Earth's crust–mantle system inferred from hafnium isotopes in chondrites

Martin Bizzarro[†], Joel A. Baker[†], Henning Haack[†], David Ulfbeck[†] & Minik Rosing^{†*}

^{*} Geological Museum, Øster Voldgade 5–7, [†] Danish Lithosphere Center, Øster Voldgade 10, 1350 Copenhagen K, Denmark

The ^{176}Lu to ^{176}Hf decay series has been widely used to understand the nature of Earth's early crust–mantle system^{1–6}. The interpretation, however, of Lu–Hf isotope data requires accurate knowledge of the radioactive decay constant of ^{176}Lu ($\lambda_{176\text{Lu}}$), as well as bulk-Earth reference parameters. A recent calibration of the $\lambda_{176\text{Lu}}$ value calls for the presence of highly unradiogenic hafnium in terrestrial zircons with ages greater than 3.9 Gyr, implying widespread continental crust extraction from an isotopically enriched mantle source more than 4.3 Gyr ago⁷, but does not provide evidence for a complementary depleted mantle reservoir. Here we report Lu–Hf isotope measurements of different Solar System objects including chondrites and basaltic eucrites. The chondrites define a Lu–Hf isochron with an initial $^{176}\text{Hf}/^{177}\text{Hf}$ ratio of 0.279628 ± 0.000047 , corresponding to $\lambda_{176\text{Lu}} = 1.983 \pm 0.033 \times 10^{-11} \text{ yr}^{-1}$ using an age of 4.56 Gyr for the chondrite-forming event. This $\lambda_{176\text{Lu}}$ value indicates that Earth's oldest minerals were derived from melts of a mantle source with a time-integrated history of depletion rather than enrichment⁷. The depletion event must have occurred no later than 320 Myr after planetary accretion, consistent with timing inferred from extinct radionuclides^{8–10}.

The main stage of Earth's accretion was completed approximately 30 Myr (refs 11, 12) after the condensation of the first solids from the early solar nebula $4,567.2 \pm 0.6$ Myr ago¹³. However, the exact timing of crust–mantle separation and onset of convection in the Earth's mantle remains poorly constrained, a critical consideration for the understanding of early planetary differentiation processes. Formation of either oceanic or continental crust results in chemical depletion, and therefore isotopic fractionation, of the residual mantle with time. Moreover, when crust is extracted, it records the isotopic composition of its source reservoir and thus provides information regarding the time-integrated history of either depletion or enrichment of its source. The Lu–Hf isotopic system, with a half-life of about 37 Gyr (ref. 7), is both a powerful chronometer and petrogenetic tool, allowing precise dating of geological materials as well as the identification of ancient Lu/Hf fractionation in planetary geochemical reservoirs.

Intrinsic to application of such isotopic tracers to the study of the early Earth is knowledge of chondritic (bulk Earth) reference values

for the Lu–Hf system (see Fig. 1) and the precise value of the decay constant for the parent isotope (^{176}Lu). A recent calibration⁷ of $\lambda_{176\text{Lu}}$ yielded a value of $1.865 \pm 0.015 \times 10^{-11} \text{ yr}^{-1}$, approximately 4% slower than previous estimates, but in agreement with two recent decay-counting experiments^{14,15}. This value has remained controversial, because the Hf isotopic composition of Earth's oldest minerals (≥ 4 -Gyr-old detrital zircons from the Jack Hills meta-sediments, Australia⁵), when back-calculated with this $\lambda_{176\text{Lu}}$ value, indicate derivation from isotopically enriched source reservoirs ($\varepsilon_{\text{Hf}} = -2$ to -6 between 3.97 and 4.14 Gyr ago; the ε_{Hf} value is the deviation of the $^{176}\text{Hf}/^{177}\text{Hf}$ ratio of a sample from the chondritic uniform reservoir—CHUR—in parts per 10^4). These unradiogenic Hf isotope signatures require that Earth had developed a first persistent crust of felsic nature by 4.3 Gyr ago, and questions the existence of a possible early mafic proto-crust¹⁶ because model ages for mafic protoliths are ~ 200 Myr older, that is, during the early stage of planetary accretion. In contrast, the evidence for a complementary reservoir of high $^{176}\text{Hf}/^{177}\text{Hf}$ values that would account for widespread pre-3.8-Gyr crustal production is lacking, because younger (~ 3.5 – 3.8 Gyr) juvenile lithologies from Greenland, South Africa and Australia seem to be derived from an unfractionated mantle source region (Fig. 2a), a feature that is not well reconciled with Nd isotope studies of early Archaean lithologies¹⁷.

Previous efforts to constrain the initial Hf isotope composition of the Solar System were either plagued by large analytical uncertainties^{18,19}, or based on the average present-day value for a number of carbonaceous and ordinary chondrites, which failed to define an isochron relationship²⁰. The latter approach is problematical, because it is highly dependent on precise knowledge of both the $^{176}\text{Lu}/^{177}\text{Hf}$ ratio for CHUR and the $\lambda_{176\text{Lu}}$ value. In order to accurately re-determine the initial $^{176}\text{Hf}/^{177}\text{Hf}$ ratio of the Solar

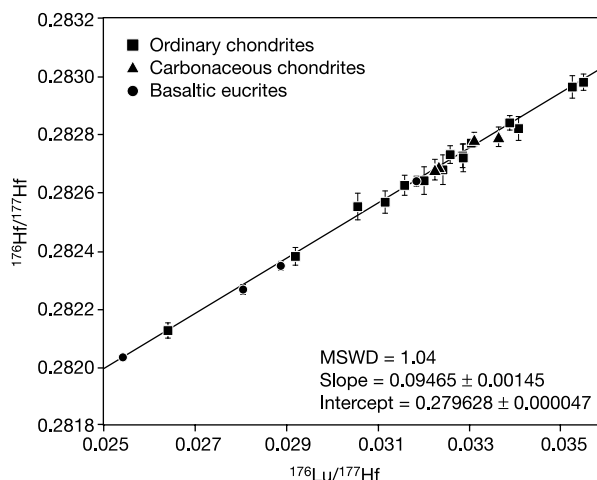


Figure 1 Lu–Hf isochron for chondrites analysed in this study. We interpret the intercept of the regression to represent the initial $^{176}\text{Hf}/^{177}\text{Hf}$ ratio of the Solar System. MSWD, mean square of weighted deviations. Errors on the $^{176}\text{Lu}/^{177}\text{Hf}$ ratios are 0.2% (2 s.d.) and are smaller than the plotting symbols. Error bars on the $^{176}\text{Hf}/^{177}\text{Hf}$ determinations represent 2 s.d. reproducibility, based on a relationship between in-run precision and external reproducibility of an in-house Hf standard. Excluding the carbonaceous chondrites yields a slope and intercept of 0.09465 ± 0.00147 and 0.279629 ± 0.000047 , respectively. Likewise, the L chondrites alone yield a slope and intercept of 0.09503 ± 0.00154 and 0.279619 ± 0.000049 . A recent comprehensive Lu–Hf study of basaltic eucrites yielded a slope and intercept of 0.0937 ± 0.0016 and 0.279638 ± 0.000046 , identical to the values determined here for chondrites (ref. 27; the eucrite Caldera, an impact melt, was excluded from the regression because Sm–Nd systematics indicate post-crystallization disturbance). All regressions were performed using isochron software available at ftp.ens-lyon.fr/pub/users/GEOL/albarede (written in Matlab, ref. 27). Data shown in this figure, together with details of the experimental methods used in their determination, are available as Supplementary Information.

System and the $\lambda_{176\text{Lu}}$ value, we have measured Lu–Hf concentrations and Hf isotope ratios on a number of Solar System objects by multiple collector inductively coupled plasma mass spectrometry (MC-ICP-MS). The materials analysed include a variety of carbonaceous and ordinary chondrites, as well as a small number of basaltic eucrites. Chondrites are among the most primitive objects derived from the solar nebula, and are considered to represent our best estimate of the average composition of the Solar System, particularly with respect to the abundance of refractory elements such as Lu and Hf (ref. 21). Basaltic eucrites are believed to have been generated by large-degree melting of a chondritic source and formed within a few million years following condensation of the first solids from the solar nebula^{22–24}. Similarly, the time interval for the formation of the chondrite parent bodies is within 10 Myr of the birth of the Solar System^{25,26}.

Our new Lu and Hf data are shown in Fig. 1. The ordinary chondrites have a relatively large range of $^{176}\text{Lu}/^{177}\text{Hf}$ and $^{176}\text{Hf}/^{177}\text{Hf}$ ratios. The large range of $^{176}\text{Lu}/^{177}\text{Hf}$ values may be potentially related to variations in phosphate abundances in equilibrated chondrites, which is supported by the observation that most of the observed variation is defined by this type of material. Basaltic eucrites show a range of $^{176}\text{Lu}/^{177}\text{Hf}$ and $^{176}\text{Hf}/^{177}\text{Hf}$ ratios comparable to the ordinary chondrites, but lie towards the

lower end of the array, a feature that is due to the partial melting processes that produced the eucrites. In contrast, the carbonaceous chondrite materials have a more restricted range of $^{176}\text{Lu}/^{177}\text{Hf}$ and $^{176}\text{Hf}/^{177}\text{Hf}$ ratios, that is, 0.0324–0.0337 and 0.28268–0.28279. No obvious correlation exists between the $^{176}\text{Hf}/^{177}\text{Hf}$ ratio and the Hf concentration for either chondrites or eucrites, suggesting that the isochron (Fig. 1) is not a fictitious result of mixing different isotopic components (for example, pre-solar versus solar).

Nineteen analyses of ordinary and carbonaceous chondrites (including replicates) define a statistically significant Lu–Hf isochron with a slope of 0.09465 ± 0.00145 and intercept of 0.279628 ± 0.000047 (Fig. 1). The slope and intercept derived from individual regressions of either the ordinary chondrites or the L type alone are identical within analytical uncertainty (see Fig. 1 legend). A notable feature of Fig. 1 is that the four basaltic eucrites align on the same array as the chondrites. As such, the combined groups of meteorites also define a statistically significant isochron with a slope of 0.09462 ± 0.00068 and intercept of 0.279627 ± 0.000020 , identical to the values derived from the chondrites alone. Moreover, a recent comprehensive Lu–Hf study²⁷ of basaltic eucrites also yielded a slope and intercept identical (within analytical uncertainty) with those determined in this study (see Fig. 1 legend). No resolvable differences in either age or initial $^{176}\text{Hf}/^{177}\text{Hf}$ ratio can thus be detected between the basaltic eucrites and chondrites analysed, indicating that the chondrite and eucrite parent bodies accreted from material of homogeneous Hf isotope composition.

Using a mean age of 4.56 Gyr for the chondrite-forming event, we derive a value for $\lambda_{176\text{Lu}}$ of $1.983 \pm 0.033 \times 10^{-11} \text{ yr}^{-1}$ from the regression of the chondrite meteorites. Even an age difference of 20 Myr in the accretion timescales of the parent bodies for the different classes of meteorites analysed results only in a shift of <0.4% in the slope of the regression—four times less than the uncertainty on the slope of the regression and clearly negligible. The presence of a minor (<4%) decay branch of ^{176}Lu to ^{176}Yb by electron capture²⁸ introduces an overestimation of $\lambda_{176\text{Lu}}$ by <0.18% for 4.56-Gyr-old samples⁷, and we have thus ignored this process, as it would result in a bias that is again insignificant relative to the uncertainty of the $\lambda_{176\text{Lu}}$ value determined here.

Our $\lambda_{176\text{Lu}}$ value of $1.983 \pm 0.033 \times 10^{-11} \text{ yr}^{-1}$ is nearly 6% faster than the recent calibration determined against the U–Pb system⁷, but within analytical error of the values determined from some published Lu–Hf eucrite isochrons^{18,27}. The consistent offset towards faster $\lambda_{176\text{Lu}}$ values for these calibrations determined against extraterrestrial objects of known age as compared to terrestrial material is problematical (both underpinned by U–Pb dating). The calibration of ref. 7 relies on Lu–Hf analysis of exotic rare-earth-enriched minerals such as gadolinite, xenotime and apatite. A potential limitation to the previous approach is that the closure temperature of the Lu–Hf system in these large minerals is poorly constrained. A significantly lower closure temperature for the Lu–Hf system compared to the U–Pb system, such as recently suggested for metamorphic apatite²⁹, may potentially resolve the discrepancy between the Lu decay constant calibrations based on extraterrestrial material as compared to terrestrial rocks and/or minerals.

The initial $^{176}\text{Hf}/^{177}\text{Hf}$ ratio of the Solar System determined here is lower by 5.4 ϵ_{Hf} units than previous estimates based on eucrite meteorites¹⁹. In contrast, the present-day bulk-Earth $^{176}\text{Hf}/^{177}\text{Hf}$ ratio calculated with the $\lambda_{176\text{Lu}}$ value and initial $^{176}\text{Hf}/^{177}\text{Hf}$ ratio of the Solar System derived from this study, and a $^{176}\text{Lu}/^{177}\text{Hf}$ ratio of 0.0332, is identical within analytical uncertainty to the value in ref. 20. Our results thus support the observation that the present-day Hf–Nd bulk-Earth isotopic value is slightly displaced towards lower $^{176}\text{Hf}/^{177}\text{Hf}$ ratios ($\sim 3\epsilon_{\text{Hf}}$) compared to the modern mantle array, which requires the presence of a hidden reservoir in Earth's mantle^{20,30}. With these parameters and our $\lambda_{176\text{Lu}}$ value, we re-evaluate the extent and timing of depletion of the terrestrial

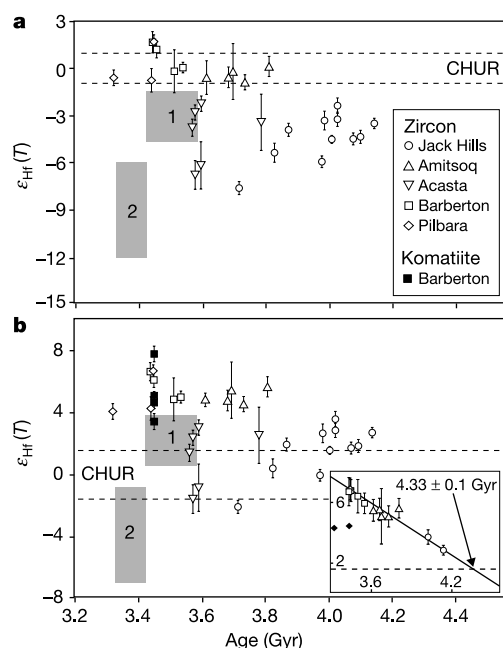


Figure 2 Hf isotope evolution diagrams for early Archaean igneous and detrital zircons and komatiites. Data taken from refs 5, 6 and 31. **a**, Calculated with the $\lambda_{176\text{Lu}}$ value of ref. 7 and the CHUR parameters of ref. 20. **b**, Calculated with the $\lambda_{176\text{Lu}}$ value and CHUR parameters from this study. Dashed lines represent the uncertainty of the CHUR value from ref. 20 (**a**) and this study (**b**). Data for the Jack Hills are ϵ_{Hf} and $^{207}\text{Pb}/^{206}\text{Pb}$ ages of individual zircons⁵. Grey boxes 1 and 2 represent distinct zircon populations from the Jack Hills. Other data correspond to the weighted average ϵ_{Hf} of zircon populations from distinct rock samples calculated with the corresponding U–Pb age (upper intercept of concordia diagram⁶). We have also included Hf data (weighted average) for a 3.81-Gyr-old Amitsoq gneiss sample³ for which whole-rock and zircon data indicate closed-system behaviour following crystallization. Inset of **b** shows ϵ_{Hf} and $^{207}\text{Pb}/^{206}\text{Pb}$ ages for most radiogenic zircons of various populations. A best fit through the data yields a minimum depletion age for Earth's mantle of 4.33 ± 0.1 Gyr, considering the uncertainty of the CHUR value. Regression parameters are: slope, -5.74 ± 0.67 ; intercept, 26.5 ± 2.6 ; and MSWD, 1.16. Zircons from the Acasta gneiss complex were not included because of evidence for ancient lead loss⁶. In addition, one sample from Barberton was excluded because all zircons were highly discordant ($\geq 10\%$). Two samples from the Pilbara⁶ plot clearly below the trend (solid diamonds) and are not included in the regression. Error bars are 2 s.d. CHUR, chondritic uniform reservoir. T, time.

mantle. Figure 2b illustrates Hf isotope data for >3.5-Gyr-old single zircons^{5,6} and deep-seated 3.45-Gyr komatiites³¹, plotted as a function of age. Only zircons for which integrated U–Pb and Hf isotope data are available were selected, thus avoiding the possibility that the original isotopic signature has been perturbed by younger geological events.

The most important result emerging from our re-calibration of the $\lambda_{176\text{Lu}}$ value is that most >4-Gyr-old zircons have $\varepsilon_{\text{Hf}} > 0$ (within analytical uncertainty), indicating that these minerals were ultimately derived from a mantle source region with a significant record of depletion. Similarly, younger zircons (<3.8 Gyr) from Greenland, Canada, Australia and South Africa, as well as 3.45-Gyr komatiites from South Africa, were extracted from depleted source regions, in accordance with the extent of fractionation recorded by the >4-Gyr grains. The felsic lithologies from which the zircons crystallized were obviously not directly derived from melting of a mantle source, but involved further differentiation processes and/or re-melting of older perhaps more mafic lithologies, which may have involved a significant time interval. As such, the Hf isotopic composition recorded by these zircons is considered a minimum estimate of the degree of depletion. We thus use the most radiogenic composition recorded by individual grains of the different age groups in Fig. 2b, in order to estimate a $^{176}\text{Lu}/^{177}\text{Hf}$ ratio of 0.043 for pre-3.45-Gyr terrestrial depleted mantle. This corresponds to $f_{\text{Lu/Hf}} \approx 0.30$ (where $f_{\text{Lu/Hf}} = [^{176}\text{Lu}/^{177}\text{Hf}]_{\text{mantle}}/[^{176}\text{Lu}/^{177}\text{Hf}]_{\text{CHUR}}$) for the pre-3.45-Gyr mantle compared to $f_{\text{Lu/Hf}} \approx 0.15$ for the modern mantle as sampled by mid-ocean ridge basalts today, indicating a significantly greater depletion of the early Archaean mantle as suggested by a number of studies^{3,4,8,16,17}.

On the basis of these fractionations, we estimate that Earth's mantle differentiated from a uniform chondritic reservoir ≈ 4.23 Gyr ago (considering the uncertainty of ± 1.7 ε_{Hf} of the new CHUR value, and the error of the slope of the regression; Fig. 2b inset). Differentiation of Earth at ~ 4.3 Gyr ago would have occurred during the life span of the now extinct ^{146}Sm nuclide and, as such, resolvable ^{142}Nd excess should be present in early Archaean lithologies before effective mantle homogenization through convective stirring. This is consistent with recent ^{146}Sm – ^{142}Nd analyses of meta-basalts from Isua (Greenland)^{9,10}, confirming the original discovery of a ^{142}Nd anomaly⁸. The magnitude of the reported ^{142}Nd excess (~ 15 – 30 p.p.m.) requires differentiation of Earth's mantle no later than 4.3 Gyr ago⁸, in excellent agreement with our results and interpretation. □

Received 4 October 2002; accepted 6 January 2003; doi:10.1038/nature01421.

1. Stevenson, K. K. & Patchett, P. J. Implications for the evolution of continental-crust from Hf-isotope systematics of Archaean detrital zircons. *Geochim. Cosmochim. Acta* **54**, 1683–1697 (1990).
2. Vervoort, J. D., Patchett, P. J., Gehrels, G. E. & Nutman, A. P. Constraints on early Earth differentiation from hafnium and neodymium isotopes. *Nature* **379**, 624–627 (1996).
3. Vervoort, J. D. & Blichert-Toft, J. Evolution of the depleted mantle: Hf isotope evidence from juvenile rocks through time. *Geochim. Cosmochim. Acta* **63**, 533–556 (1999).
4. Albarède, F., Blichert-Toft, J., Vervoort, J. D., Gleason, J. G. & Rosing, M. Hf–Nd evidence for a transient dynamic regime in early terrestrial mantle. *Nature* **404**, 488–490 (2000).
5. Amelin, Y., Lee, D.-C., Halliday, A. N. & Pidgeon, R. T. Nature of the Earth's earliest crust from hafnium isotopes in single detrital zircons. *Nature* **399**, 252–255 (1999).
6. Amelin, Y., Lee, D.-C. & Halliday, A. N. Early-middle Archaean crustal evolution deduced from Lu–Hf and U–Pb isotopic studies of single grain zircons. *Geochim. Cosmochim. Acta* **64**, 4205–4225 (2000).
7. Scherer, E., Münker, C. & Mezger, K. Calibrating the Lu–Hf clock. *Science* **293**, 683–686 (2001).
8. Harper, C. L. Jr & Jacobsen, S. B. Evidence from coupled ^{147}Sm – ^{143}Nd and ^{146}Sm – ^{142}Nd systematics for very early (4.5-Gyr) differentiation of the Earth's mantle. *Nature* **360**, 728–732 (1992).
9. Boyet, M., Albarède, F., Télouk, P. & Rosing, M. ^{142}Nd anomaly confirmed at Isua. *Geochim. Cosmochim. Acta* **66**, Abst. A99 (2002).
10. Caro, G., Bourdon, B., Birck, J.-L. & Moorbath, S. $^{142}\text{Nd}/^{144}\text{Nd}$ precise determination in early Archaean rocks. *Geochim. Cosmochim. Acta* **66**, Abst. A120 (2002).
11. Yin, Q. et al. A short timescale for terrestrial planet formation from Hf–W chronometry of meteorites. *Nature* **418**, 949–952 (2002).
12. Kleine, T., Münker, C., Mezger, K. & Palme, H. Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf–W chronometry. *Nature* **418**, 952–956 (2002).
13. Amelin, Y., Krot, A. N., Hutcheon, I. D. & Ulyanov, A. A. Lead isotopic ages of chondrules and calcium–aluminium-rich inclusions. *Science* **297**, 1678–1683 (2002).

14. Dalmasso, J., Barci-Funel, G. & Ardisson, G. J. Reinvestigation of the decay of the long-lived odd-odd ^{176}Lu nucleus. *Appl. Radiat. Isot.* **43**, 69–76 (1992).
15. Nir-el, Y. & Lavi, N. Measurement of the half-life of ^{176}Lu . *Appl. Radiat. Isot.* **49**, 1653–1655 (1998).
16. Chase, C. G. & Patchett, P. J. Stored mafic/ultramafic crust and early Archaean mantle depletion. *Earth Planet. Sci. Lett.* **91**, 66–72 (1988).
17. Bennett, V. C., Nutman, A. P. & McCulloch, M. T. Nd isotopic evidence for transient, highly depleted mantle reservoirs in the early history of the Earth. *Earth Planet. Sci. Lett.* **119**, 299–317 (1993).
18. Patchett, P. J. & Tatsumoto, M. Lu–Hf total-rock isochron for the eucrite meteorites. *Nature* **288**, 571–574 (1980).
19. Tatsumoto, M., Unruh, D. M. & Patchett, P. J. U–Pb and Lu–Hf systematics of Antarctic meteorites. *Proc. 6th Symp. on Antarctic meteorites: Mem. Nat. Inst. Polar Res. (Tokyo) Spec. Iss.* **20**, 237–249 (1981).
20. Blichert-Toft, J. & Albarède, F. The Lu–Hf isotope geochemistry of chondrites and the evolution of the mantle–crust system. *Earth Planet. Sci. Lett.* **148**, 243–258 (1997).
21. Alexander, C. M. O'D., Boss, A. P. & Carlson, R. W. The early evolution of the inner solar system: A meteoritic perspective. *Science* **293**, 64–68 (2001).
22. Lugmair, G. W. & Shukolyukov, A. Early solar system timescales according to ^{53}Mn – ^{53}Cr systematics. *Geochim. Cosmochim. Acta* **62**, 2863–2886 (1998).
23. Lugmair, G. W. Sm–Nd ages: a new dating method. *Meteoritics* **9**, 369 (1974).
24. Lugmair, G. W. & Scheinin, N. B. Sm–Nd systematics of the Stannern meteorite. *Meteoritics* **10**, 447–448 (1975).
25. Jacobsen, S. B. & Wasserburg, G. J. Sm–Nd isotopic evolution of chondrites. *Earth Planet. Sci. Lett.* **50**, 139–155 (1980).
26. Göpel, C., Manhès, G. & Allègre, C. U–Pb systematics from equilibrated ordinary chondrites. *Earth Planet. Sci. Lett.* **121**, 153–171 (1994).
27. Blichert-Toft, J., Boyet, M., Télouk, P. & Albarède, F. ^{147}Sm – ^{143}Nd and ^{176}Lu – ^{176}Hf in eucrites and the differentiation of the HED parent body. *Earth Planet. Sci. Lett.* **204**, 167–181 (2002).
28. Dixon, D., McNair, A. & Curran, S. C. The natural radioactivity of lutetium. *Phil. Mag.* **45**, 683–694 (1954).
29. Barfod, G. H. A New Lu–Hf Separation Technique for Phosphates and its Application to Apatite Geochronology. Thesis, Univ. Copenhagen (2002).
30. Bizzarro, M., Simonetti, A., Stevenson, R. K. & David, J. Hf isotope evidence for a hidden mantle reservoir. *Geology* **30**, 771–774 (2002).
31. Blichert-Toft, J. & Arndt, N. T. Hf isotope compositions of komatiites. *Earth Planet. Sci. Lett.* **171**, 439–451 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank F. Albarède and H. C. Larsen for comments on an earlier version of the manuscript, J. Blichert-Toft for access to unpublished data, Y. Amelin for discussion, and T. Wright for help in the Danish Lithosphere Centre MC-ICP-MS laboratory. This project was supported by the Danish Lithosphere Centre (funded by the Danish National Science Foundation). M.B. was supported by an NSERC postdoctoral fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.B. (e-mail: mbi@dlc.ku.dk).

Patterns and processes in reef fish diversity

Camilo Mora*, Paul M. Chittaro*, Peter F. Sale*, Jacob P. Kritzer* & Stuart A. Ludsin*†

* Department of Biology, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4 Canada

A central aim of ecology is to explain the heterogeneous distribution of biodiversity on earth. As expectations of diversity loss grow^{1–5}, this understanding is also critical for effective management and conservation. Although explanations for biodiversity patterns are still a matter for intense debate⁵, they have often been considered to be scale-dependent^{6,7}. At large geographical scales, biogeographers have suggested that variation in species richness results from factors such as area, temperature, environmental stability, and geological processes, among many

† Present address: National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan 48105-2945, USA.

others^{5,7–14}. From the species pools generated by these large-scale processes, community ecologists have suggested that local-scale assembly of communities is achieved through processes such as competition, predation, recruitment, disturbances and immigration^{5–8,15,16}. Here we analyse hypotheses on speciation and dispersal for reef fish from the Indian and Pacific oceans and show how dispersal from a major centre of origination can simultaneously account for both large-scale gradients in species richness and the structure of local communities.

Using a geographical database of species distributions (see Methods), we determined large-scale patterns in species richness for reef fish in the Indian and Pacific oceans. From a centre of high diversity in the Indonesian and Philippine region (IPR), species richness decreases steadily along both latitudinal and longitudinal axes (Fig. 1a, b). Although these patterns are similar in shape to those that one would expect from the mid-domain effect (the pattern expected if species ranges were randomly distributed between geographical boundaries¹⁰), the overall patterns, and particularly the high number of species in the IPR, result from a nonrandom distribution of species (Fig. 1a, b).

Three major, yet different, hypotheses invoking speciation and dispersal have been suggested to explain these large-scale patterns. The Centre-of-Origin hypothesis suggests that the IPR is a major centre of speciation from which species disperse to marginal locations^{9,12,17}. The Centre-of-Overlap hypothesis proposes that the high diversity in the IPR is due to the overlapping of faunas from several biogeographic provinces^{9,12,17}. Finally, the Centre-of-Accumulation hypothesis states that speciation occurs in several areas peripheral to the IPR and that species extend their ranges to the IPR by prevailing currents^{9,12,17}. A variant to the Centre-of-Accumulation hypothesis holds that after extending into the IPR, many species have suffered range reductions through the loss of populations marginal to the IPR¹². Whereas the Centre-of-Overlap hypothesis predicts dispersal in all directions within the biogeographic provinces, the Centre-of-Accumulation hypothesis (including its variant) predicts unidirectional dispersal towards the IPR.

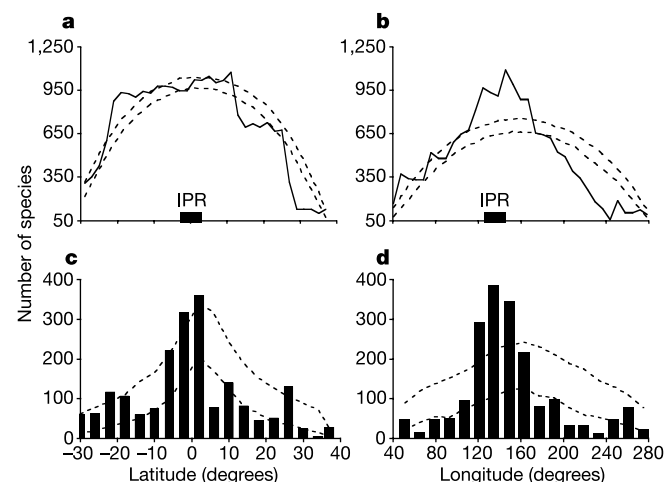


Figure 1 Geographical patterns in reef fish biodiversity in the Indian and Pacific oceans. IPR, Indonesian and Philippine region. Latitudinal (a) and longitudinal (b) clines (solid lines) were defined as the number of species whose geographical ranges included a point in latitude or longitude, respectively. Distributions of mid-latitudinal (c) and mid-longitudinal (d) ranges (filled bars) are also shown. The effects of geographic constraints on such patterns (the mid-domain effect¹⁰) were tested by running a null model in which the ranges (for a and b) and mid-ranges (for c and d) were randomly allocated between boundaries. These boundaries were the 'hard' limits imposed by the coasts of Africa and America in longitude and the 'soft' limits implied by the 37° N and 32° S latitude where tropical organisms show striking reductions in species richness²⁰. Dotted lines correspond to the maximum and minimum values after running the model 1,000 times.

The Centre-of-Overlap and the Centre-of-Accumulation hypotheses state that only the tails of species' ranges extend into the IPR. Consequently, most species should have their range midpoints marginal to the IPR, resulting in bimodal or multi-modal distributions. However, plots of species mid-ranges (both longitudinal and latitudinal) show nonrandom unimodal distributions with peaks coinciding with the geographical position of the IPR (Fig. 1c, d). These distributions rule out these two hypotheses and provide support for the Centre-of-Origin hypothesis. They also support, to some extent, the variant of the Centre-of-Accumulation hypothesis because range reductions through loss of peripheral populations would shift mid-ranges towards the IPR, and, if extensive, this could result in a unimodal distribution of mid-ranges in the vicinity of the IPR.

The Centre-of-Origin and the Centre-of-Accumulation (including its variant) hypotheses predict different places of species origination. Whereas the Centre-of-Origin hypothesis predicts speciation within the IPR, the Centre-of-Accumulation hypothesis predicts speciation in locations marginal to the IPR. To discern between these predictions, we analysed the geographic pattern in reef fish endemism. We assume that centres of endemism contain a preponderance of recently derived species that are yet to expand their ranges (neo-endemics) and thus provide insights into areas where species are most likely to originate. It is reasonable to expect that palaeo-endemic species, if present, will represent only a small proportion of endemics. Palaeo-endemic extinction rates should be high because of their small ranges and longer exposure to factors causing extinction. The IPR stands out as the major centre of endemism in the Indian and Pacific oceans (Fig. 2), as has been previously reported for fish as well as for other taxa⁴. This result supports the IPR as a major centre of speciation and confirms the expectation of the Centre-of-Origin hypothesis. In addition, Springer¹¹ has argued that large-scale extinctions have not occurred marginal to the IPR, casting doubt on the variant of the Centre-of-Accumulation hypothesis.

(We note that Hughes *et al.*¹⁸ have recently reported low levels of endemism for the IPR. We believe their result, which is in contrast to our results and previous studies, is partially a consequence of the manner in which they sub-sampled their database. Specifically, they excluded sites less than 400 km apart, which means that many proximal sites within the IPR might be excluded and perceived endemism within the region would be artificially reduced.)

Our analysis also demonstrates that other minor centres of endemism exist, which might contribute to regional species pools. Most of these minor centres are, however, geographically isolated or in places where current direction is predominately from tropical to temperate latitudes⁴ (Fig. 2). Thus, species generated in these outlying centres of origin probably would not contribute significantly to species pools in other communities. By contrast, the IPR occurs in a highly interconnected area that may facilitate species migration or dispersal to marginal locations. That the IPR harbours

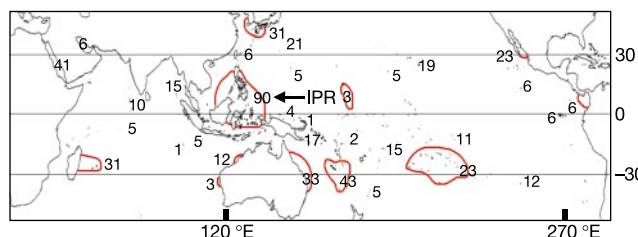


Figure 2 Geographical pattern of reef fish endemism in the Indian and Pacific oceans. Endemic species were defined as those species restricted to a single location in the database. Numbers in the graph indicate the number of endemic species at each location. To simplify the figure, we combined endemic species of nearby locations into a single number. These locations are enclosed with a red line.

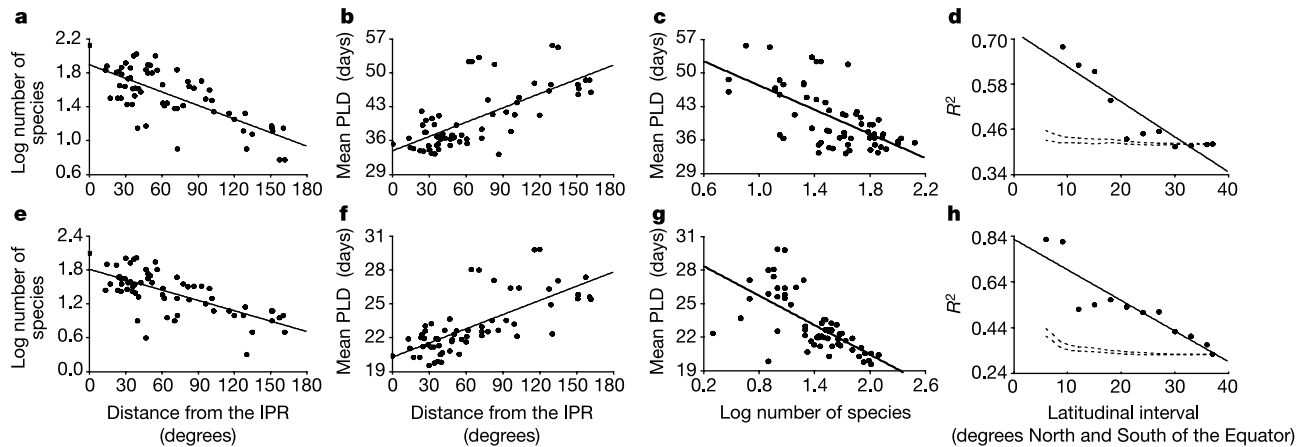


Figure 3 Pairwise comparisons of species richness, pelagic larval duration and distance from the Indonesian and Philippine region. Data is shown for Labridae (**a**: $r = -0.73$, $P < 0.001$; **b**: $r = 0.70$, $n = 62$, $P < 0.001$; **c**: $r = -0.64$, $n = 62$, $P < 0.001$) and Pomacentridae (**e**: $r = -0.69$, $P < 0.001$; **f**: $r = 0.72$, $P < 0.001$; **g**: $r = -0.67$, $P < 0.001$). The sample size in all correlations, except **b** and **c**, was 63 locations between 29° N and 29° S. Panels **d** and **h** show the coefficients of determination for regression analyses between species richness (dependent variable) and mean pelagic

larval duration (PLD) (independent variable) for 12 latitudinal intervals (**d**: Labridae: $r = -0.92$, $P < 0.001$; and **h**: Pomacentridae: $r = -0.90$, $P < 0.001$). We tested the extent to which the latitudinal pattern in the coefficient of determination was a statistical artefact by simulating the regressions for each latitudinal interval, using the same number of locations, but randomly selecting these from all latitudes. Dotted lines are 95% limits of confidence for coefficients of determination of 1,000 regression analyses for each of the 12 latitudinal intervals.

among the highest number of islands per unit of geographical area makes it a place where allopatric speciation might be frequent, especially when considering patterns of recent geological sea level change. Whatever the ultimate causal factor, the IPR seems to be a centre of species origination and a source of species for reef fish communities in the Indian and Pacific oceans.

The notion that reef fishes originate in the IPR and disperse to marginal communities leads to two testable predictions about patterns in reef fish biodiversity: (1) both the longitudinal and latitudinal gradients of species richness, as one moves away from the IPR, are due to variation in dispersal abilities; and (2) this process will supply the species structuring communities marginal to the IPR. To test the first prediction, we correlated data on species richness for all locations analysed with the mean pelagic larval duration (PLD) of the species inhabiting those locations. This analysis was restricted to labrids and pomacentrids, the only families for which extensive information on PLD exists (see Methods). We found that species richness in both families was negatively correlated with distance from the IPR (Fig. 3a, e), and that mean PLD for these families increased with increasing distance from the IPR¹⁹ (Fig. 3b, f). Consequently, species richness was negatively correlated with mean PLD at local sites (Fig. 3c, g). These results indicate that species richness declines as one moves away from the IPR, probably owing to limitations in dispersal. Using multiple regression analysis, we found that 57% and 60% of the geographical variation in species richness of Labridae and Pomacentridae, respectively, was explained by PLD and distance from the IPR.

Interestingly, we also found that communities at high latitudes were usually outliers in the relationship between PLD and species richness. The geographical variation in species richness that is explained by PLD increases when the latitudinal scope of the analysis is reduced (Fig. 3d, h). We presume that shallow-water habitats, which are poorly interconnected in longitude owing to deep-water gaps between islands and coasts, but continuous in latitude owing to the margins of the continents, make PLD more critical for expansion in longitude than for expansion in latitude. Additionally, sharp barriers (for example, current and temperature shifts) may limit species expansion at high latitudes regardless of PLD²⁰. This suggests that PLD gains additional importance for species distribution and richness where geographical isolation is predominant, and that, owing to the shape of the Indian and Pacific

oceans, the extent to which different processes regulate local communities may vary geographically. Assembly of local communities at lower latitudes is probably determined by the distance to, and the abilities of species to disperse across, open water from the IPR, whereas extreme abiotic factors may cause local extinctions at higher latitudes regardless of dispersal capabilities.

If the IPR is a centre of origin and a source of species to reef fish communities in the Indian and Pacific oceans and if local processes affect local diversity only minimally, then reef fish communities throughout this region should have a large number of species in common with the IPR. Indeed, we found that about 86% of the species comprising reef fish assemblages in the Indian and Pacific oceans were species present in the IPR, a result that differs from expectations based on random assembly (Fig. 4). Moreover, local speciation seems to have a minor role in structuring marginal communities given that only around 2% of the species in any community were endemics (Fig. 4). These results suggest that speciation in and dispersal from the IPR play a major role in assembling communities in the Indian and Pacific oceans. That is, dispersal governs which species from the IPR are capable of reaching communities at marginal locations. The fact that the total number of species present in any location is almost equal to the number of

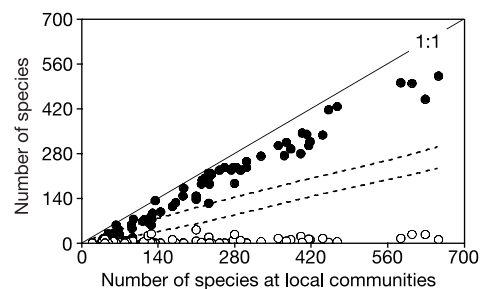


Figure 4 Contribution of IPR and endemic species to local reef fish assemblages in the Indian and Pacific oceans. Filled circles, IPR species; open circles, endemic species. We assessed the extent to which the number of IPR species at local communities is due to chance by determining the number of IPR species in communities (of the same size as observed) randomly generated from the total species pool (species selected with equal probability and without replacement). The upper and lower limits of the number of IPR species in 1,000 iterations for each community are shown as broken lines.

species that it shares with the IPR suggests that no location contributes as much to the overall alpha diversity of the Indian and Pacific oceans as does the IPR.

The distribution of biodiversity on Earth can be described in terms of a few well documented, and intriguing, small- to large-scale patterns. Our findings recognize a major link between the evolutionary processes regulating these patterns. That is, the processes of speciation, extinction and dispersal that yield large-scale patterns of species richness also seem to determine which species are present within local assemblages. Given the importance of the IPR to the overall structure of reef fish assemblages in the Indian and Pacific oceans, it should certainly be a target for strategic management and protection. □

Methods

Analyses were based on the presence or absence of 1,970 reef fish species in 70 locations in the Indian and Pacific oceans. These species belong to the families Labridae, Pomacentridae, Serranidae, Blenniidae, Apogonidae, Chaetodontidae, Acanthuridae, Scaridae, Holocentridae, Lutjanidae, Pomacanthidae, Scorpaenidae and Lethrinidae. These families are among the most diverse, best-known taxonomically, and represent >70% of the total species expected in any community. Owing to the high covariation in species richness among families⁸, these families are a good indicator for the remaining diversity of species. The database includes all data from the 63 locations used in ref. 8. Data for the following locations were added: the Philippines, Madagascar, Easter Island, Cook Islands (all from www.fishbase.org), Cocos Keeling²¹; Gorgona Reef²²; and Korea (http://ricos.cnu.ac.kr/~kocofish/list/elisintro.htm). Data for the following locations were updated: Galapagos²³; Gulf of California²⁴; and Malpelo (C.M., personal observations). All species records were corrected for synonymy and other taxonomic problems using The Catalog of Fishes (http://www.calacademy.org/research/ichthyology/catalog/fishcatsearch.html). More than 300 species were duplicated in the original database as a result of synonymy, misspelled names or misallocation of species to families.

Pelagic larval durations were obtained for 95 labrid species and 116 pomacentrids. These species represent 28% of all labrid species and 42% of the pomacentrids. Data were obtained from references 19, 25–27.

Received 8 October; accepted 27 December 2002; doi:10.1038/nature01393.

1. Walther, G. R. *et al.* Ecological responses to recent climate change. *Nature* **416**, 389–395 (2002).
2. Chapin, F. S. I. *et al.* Consequences of changing diversity. *Nature* **405**, 234–242 (2000).
3. Roberts, C. M. & Hawkins, J. P. Extinction in the sea. *Trends Ecol. Evol.* **14**, 241–246 (1999).
4. Roberts, C. M. *et al.* Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280–1284 (2002).
5. Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
6. Keddy, P. & Weiher, E. in *Ecological Assembly Rules* (eds Weiher, E. & Keddy, P.) 1–20 (Cambridge Univ. Press, Cambridge, UK, 1999).
7. Cornell, H. V. & Karlson, R. H. Coral species richness: Ecological vs. biogeographical influences. *Coral Reefs* **19**, 37–49 (2000).
8. Bellwood, D. & Hughes, T. Regional-assembly rules and biodiversity of coral reefs. *Science* **292**, 1532–1535 (2001).
9. Veron, J. E. N. *Coral Reefs in Space and Time: The Biogeography and Evolution of the Scleractinia* (UNSW Press, Sydney, 1995).
10. Colwell, R. K. & Lees, D. C. The mid-domain effect: geometric constraints on the geography of species richness. *Trends Ecol. Evol.* **15**, 70–76 (2000).
11. Springer, V. G. Pacific plate biogeography, with special reference to shorefishes. *Smithson. Contrib. Zool.* **177**, 1–182 (1982).
12. Bellwood, D. R. & Wainwright, P. C. in *Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem* (ed. Sale, P. F.) 5–32 (Academic, San Diego, California, 2002).
13. Harmelin-Vivien, M. L. in *Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem* (ed. Sale, P. F.) 265–274 (Academic, San Diego, California, 2002).
14. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, UK, 1995).
15. Caley, M. J. & Schluter, D. The relationship between local and regional diversity. *Ecology* **78**, 70–80 (1997).
16. Karlson, R. H. & Cornell, H. V. Scale-dependent variation in local vs. regional effects on coral species richness. *Ecol. Monogr.* **68**, 259–274 (1998).
17. Planes, S. in *Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem* (ed. Sale, P. F.) 201–220 (Academic, San Diego, California, 2002).
18. Hughes, T. P., Bellwood, D. R. & Connolly, S. R. Biodiversity hotspots, centres of endemism, and the conservation of coral reefs. *Ecol. Lett.* **5**, 775–784 (2002).
19. Bonhomme, F. & Planes, S. Some evolutionary arguments about what maintains the pelagic interval in reef fishes. *Environ. Biol. Fish.* **59**, 365–383 (2000).
20. Gaylord, B. & Gaines, S. D. Temperature or transport: species range boundaries mediated solely by flow. *Am. Nat.* **115**, 769–789 (2000).
21. Allen, G. R. & Smith-Vaniz, W. F. Fishes of the Cocos (Keeling) Islands. *Atoll Res. Bull.* **412**, 1–21 (1994).
22. Zapata, F. & Morales, Y. A. Spatial and temporal pattern of fish diversity in a coral reef at Gorgona Island, Colombia. *Proc. 8th Int. Coral Reef Symp.* **1**, 1029–1034 (1998).
23. Grove, J. S. & Lavenberg, R. J. *The Fishes of the Galapagos Islands* (Stanford Univ. Press, Stanford, 1997).
24. Thomson, D. A., Findley, L. T. & Kerstich, A. N. *Reef Fishes of the Sea of Cortez* (Univ. Texas Press, Austin, 1999).
25. Victor, B. C. Duration of the planktonic larval stage of one hundred species of Pacific and Atlantic wrasses (family Labridae). *Mar. Biol.* **90**, 317–326 (1986).
26. Victor, B. C. & Wellington, G. M. Endemism and the pelagic larval duration of reef fishes in the

eastern Pacific Ocean. *Mar. Ecol. Prog. Ser.* **205**, 241–248 (2000).

27. Victor, B. C. Planktonic larval duration of one hundred species of Pacific and Atlantic damselfishes (Pomacentridae). *Mar. Biol.* **101**, 557–567 (1989).

Acknowledgements We thank R. H. Karlson, J. Lovett-Doust, F. Zapata, J. Ciborowski and D. Hogan for discussion and comments. The staff at the Leddy Library (University of Windsor) assisted in obtaining copies of the manuscripts required to build the database. Funding was provided by NSERC (to P.F.S.) and OGS (to C.M. and P.M.C.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.M. (e-mail: moracamilo@hotmail.com).

Mechanism of genetic exchange in American trypanosomes

Michael W. Gaunt*, Matthew Yeo*, Iain A. Frame*†, J. Russell Stothard†‡, Hernan J. Carrasco§, Martin C. Taylor*, Susana Solis Mena*, Paul Veazey*, Graham A. J. Miles*, Nidia Acosta||, Antonieta Rojas de Arias|| & Michael A. Miles*

* Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

‡ Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

§ Facultad de Medicina, Universidad Central de Venezuela, Caracas, Venezuela

|| Departamento de Medicina Tropical, Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asuncion, Asuncion, Paraguay

The kinetoplastid Protozoa are responsible for devastating diseases¹. In the Americas, *Trypanosoma cruzi* is the agent of Chagas' disease—a widespread disease transmissible from animals to humans (zoonosis)—which is transmitted by exposure to infected faeces of blood-sucking triatomine bugs². The presence of genetic exchange in *T. cruzi* and in *Leishmania* is much debated^{3,4}. Here, by producing hybrid clones, we show that *T. cruzi* has an extant capacity for genetic exchange. The mechanism is unusual and distinct from that proposed for the African trypanosome, *Trypanosoma brucei*⁵. Two biological clones⁶ of *T. cruzi* were transfected to carry different drug-resistance markers^{7,8}, and were passaged together through the entire life cycle. Six double-drug-resistant progeny clones, recovered from the mammalian stage of the life cycle, show fusion of parental genotypes, loss of alleles, homologous recombination, and uni-parental inheritance of kinetoplast maxicircle DNA. There are strong genetic parallels between these experimental hybrids and the genotypes among natural isolates of *T. cruzi*. In this instance, aneuploidy through nuclear hybridization results in recombination across far greater genetic distances than mendelian genetic exchange. This mechanism also parallels genome duplication^{9,10}.

The species *T. cruzi* is divided into two divisions (I and II) on the basis of isoenzyme phenotypes, DNA profiles, ribosomal and mini-exon DNA sequence polymorphisms, and microsatellite analysis^{11–13}. In addition, *T. cruzi* II, which predominates where Chagas' disease is more severe¹⁴, can be divided into up to five sublineages (IIa–e)^{11,15}. Population genetics has emphasized clonal propagation and the lack of genetic exchange in *T. cruzi* when natural isolates, from dispersed geographical localities, have been tested for random mendelian genetic exchange^{4,8,16}. Nevertheless, recent phylogenetic evidence suggests that *T. cruzi* IId and IIe may have an ancient hybrid origin^{3,16}. Surprisingly, meticulous quantification studies demonstrated that

† Present address: The Wellcome Trust, 183 Euston Road, London NW1 2BE, UK (I.A.F.); Department of Infectious Disease Epidemiology, Imperial College, London W2 1PG, UK (J.R.S.).

biological clones of a single *T. cruzi* strain had between 30% and 70% more DNA than a parental stock¹⁷, indicating some rapid genetic mechanism for radical change in DNA content. Our experiments reconcile these observations by providing evidence of genetic hybridization in *T. cruzi*, and through evidence that this leads to aneuploidy observed in natural populations.

Two primary isolates of *T. cruzi* I were obtained from an undisturbed enzootic cycle of transmission at a site in the Brazilian Amazon basin, where putative parental and hybrid phenotypes were circulating sympatrically⁶. A pair of putative parental biological clones, P1 and P2, were transfected, respectively, with a modified form of pTEX carrying the hygromycin-resistance gene (M.C.T., unpublished observations) and with pTEX¹⁸, which carries the neomycin (G418)-resistance gene. *Trypanosoma cruzi* P1-*hyg* and *T. cruzi* P2-*neo* were then passaged singly or together through the various life cycle stages (see Methods). Immediately after copassage, recovered organisms were placed in culture with hygromycin and G418 to select for populations resistant to both drugs. Double-drug pressure was sustained throughout biological cloning of the hybrids and all their subsequent growth stages.

Fifty Vero cell cultures infected with a mixture of P1-*hyg* and P2-*neo* yielded one population of organisms resistant to both hygromycin and G418. Six derived biological clones (1C2, 1D12, 2A2, 2C1, 2D9, 2F9) were resistant to both drugs. No double-drug-resistant populations were recovered from mixed axenic epimastigote cultures, from mixed passage through triatomine bugs, or from mixed passage in SCID mice.

We used several methods to investigate the nature of the six double-drug-resistant *T. cruzi* clones. DNA amplification (Fig. 1a) showed that the six clones contained both the *hyg* and *neo* genes. Isoenzyme analysis (Fig. 1b) and karyotype analysis (Fig. 1c) showed that the clones shared parental characters and were at least, in part, hybrids. Random amplification of polymorphic DNA (RAPD) with three of seven primers revealed characteristic sharing of bands between progeny clones and parents; controls with non-transformed P1 and P2 showed that shared bands were not attributable to the episomes (Fig. 1d). Microscopy of Giemsa-stained preparations demonstrated that the hybrid clones were not binucleate.

P1-*hyg* and P2-*neo* were screened for microsatellite repeat-length polymorphisms. For six informative loci, which distinguished P1-*hyg* from P2-*neo*, each of the six progeny clones showed all parental alleles (Fig. 1e; see also Supplementary appendix 1a, b). For four other loci, with one homozygous and one heterozygous parent,

the unique allele of the heterozygous parent was passed to all progeny. However, for locus L660, the allele of the homozygous parent (P2-*neo*) was absent from all progeny, indicating allele loss (Supplementary appendix 1a).

Allele loss was also apparent among progeny at the nuclear trypanodioxin locus (*tpn1*)¹⁹, in that none of the progeny showed the unique *tpn1* allele of the heterozygous parent P2-*neo* at position 172 base pairs (bp) (A/T) and 204 bp (A/C), with P1-*hyg* being homozygous at both positions (T and C respectively). Sequence analysis of the nuclear genes glucosephosphate isomerase (*gpi*), an intergenic region (*tcp*)²⁰ and phosphoglucosyltransferase (*pgm*) indicated fusion and revealed evidence of recombination. The parents, P1 and P2, at the *gpi*, *tcp* and *pgm* loci had a combined total of three, six and four genotypes, respectively (Supplementary appendix 2). As evidence of fusion, one progeny clone (2A2) had all but one of these genotypes. Five clones had a full complement of both parental genotypes at one or more of the three loci, but several parental genotypes were not recovered from the hybrids. Putative recombination was present at the *gpi* locus (5' 600 bp), the *tcp* intergenic region (5' 400 bp) and the *pgm* locus (380 bp) in the form of mosaic parental genotypes within amplified DNA clones derived from the biological clones of the six double-drug-resistant progeny (Methods; see also Supplementary appendix 2 (GenBank accession numbers AY227811–AY227891)). Recovery of six mosaics identifiable by different genetic markers was not compatible with slippage and polymerase chain reaction (PCR) artefacts, although the frequency of recombination events could not be ascertained.

The mitochondrial maxicircle sequence revealed a difference in the NADH dehydrogenase subunit 1 (*nd1*) gene between P1-*hyg* (A) and P2-*neo* (G) at position 102 bp (613 bp in the amplicon sequenced). Five progeny (1C2, 2A2, 2C1, 2D9, 2F9) inherited the P1-*hyg* genotype and one (1D12) inherited the P2-*neo* genotype; none of the progeny inherited both types. Parental cytochrome oxidase II sequences were identical.

Analogy between the experimental hybrids and natural populations was shown by finding evidence for three or more size-length polymorphisms at single microsatellite loci within biological clones for one field isolate of *T. cruzi* I and four field isolates of *T. cruzi* II (Supplementary appendix 3). The *T. cruzi* I (TCI) isolate showed polyploidy at half of the microsatellite loci examined (6 out of 12). Multiple microsatellite genotypes previously thought to indicate multiclonality¹³ may be explicable predominantly by polyploidy. The CL Brener strain also diverges from uniform diploidy, with triploid sections of the genome (genome project strain, J. Kelly,

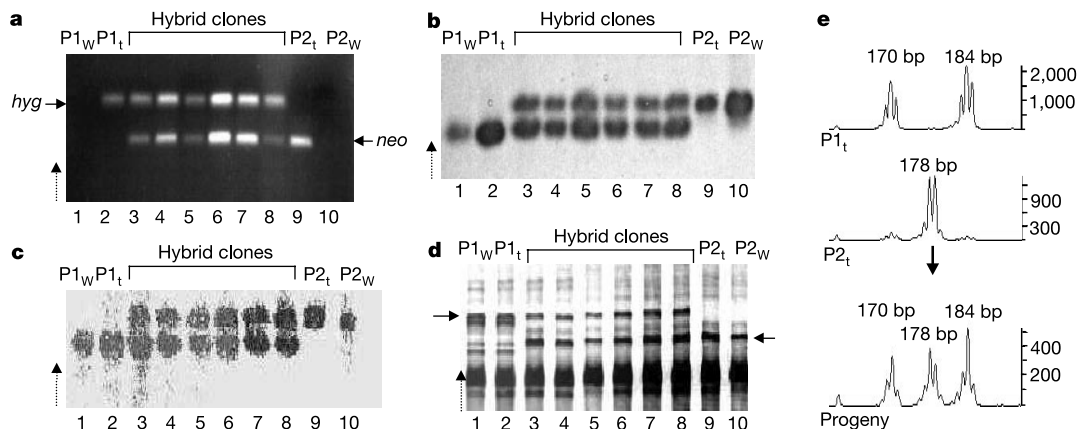
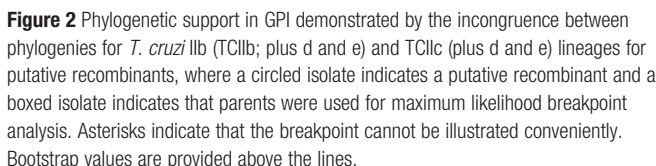


Figure 1 Hybrids of parental phenotypes and genotypes in experimentally derived double-drug-resistant biological clones of *T. cruzi*. **a**, Ethidium-stained agarose gel of multiplex polymerase chain reaction (PCR) to detect *hyg* and *neo*. Lanes (for **a-d**) are: *T. cruzi* P1 (P1_w; lane 1); transformed P1-*hyg* (P1_t; lane 2); six double-drug-resistant hybrids (lanes 3–8); transformed *T. cruzi* P2-*neo* (P2_t; lane 9); and wild-type *T. cruzi* P2 (P2_w; lane 10). **b**, Starch gel electrophoresis showing PGM phenotypes. **c**, Southern blot of contour-clamped homogeneous electric field (CHEF) gel probed to show cysteine protease karyotypes. **d**, One of three random amplification of polymorphic DNA (RAPD) profiles (smaller products omitted) showing shared bands (arrowed). **e**, One (A427) of six polyploid microsatellite loci (y = fluorescent units).

(P2_w; lane 10). **b**, Starch gel electrophoresis showing PGM phenotypes. **c**, Southern blot of contour-clamped homogeneous electric field (CHEF) gel probed to show cysteine protease karyotypes. **d**, One of three random amplification of polymorphic DNA (RAPD) profiles (smaller products omitted) showing shared bands (arrowed). **e**, One (A427) of six polyploid microsatellite loci (y = fluorescent units).



Phylogenetic analyses gave unequivocal evidence of genetic recombination among natural populations. Thus gene mosaics were seen in 7 out of 18 isolates, over 1,700 bp of total nucleotide sequence, with mosaics between TCIIb and TCIIc genotypes (six isolates) and between TCIIb and TCI (one isolate) identified using both bootstrap methods (Fig. 2; see also Supplementary appendix 4) and bootscan analysis (Fig. 3). All sites of recombination were confirmed using a maximum likelihood test of recombination breakpoints ($P \leq 0.01$ in all cases). Description of subspecific groups by phylogenetics showed that genotypes from TCIIId and TCIIe were split between groups TCIIb and TCIIc. Trees were congruent with previous phylogenetic studies (Supplementary appendix 5). There was supporting evidence of genetic recombination for the *gpi* and *tcp* loci in at least 39% of all reference strains or field isolates.

events; it is also likely to be present in *Leishmania*²⁴. The minimal divergence time separating TCIIB from the TCIIc and TCI sister lineages is estimated to be 2–9 Myr (ref. 3) using nuclear loci (8–20 Myr for a mitochondrial maxicircle locus). Thus the observed homologous recombination between these lineages spans a vast temporal divergence, based on conservative calibration points²⁵. This phenomenon provides an added dimension to the concept of genome duplication, in that it may also operate in phylogenetic groups close to the origins of eukaryotes^{9,10}. □

Biological clones of *T. cruzi* were prepared by microscopic selection of single cells and culture. The following *in vitro* or *in vivo* systems²⁶ were used for co-passage of *T. cruzi* P1-*hyg* and P2-*neo*. First, P1-*hyg* and P2-*neo* epimastigotes were mixed in equal amounts,

grown axenically *in vitro* using supplemented RPMI 1640 medium, and at intervals after the stationary phase (21 days) they were passaged into new cultures containing each drug or both drugs (hygromycin, 150 $\mu\text{g ml}^{-1}$; G418, 120 $\mu\text{g ml}^{-1}$). Second, mammalian cell (Vero) monolayers were infected with P1-*hyg* and P2-*neo* using stationary-phase, mixed axenic cultures containing epimastigotes and metacyclic trypomastigotes. Trypomastigotes from pseudocysts were recovered periodically between days 7 and 28, and grown as epimastigotes in axenic culture for drug sensitivity tests. Third, triatomine bugs were membrane fed on mouse blood containing P1-*hyg* and P2-*neo* trypomastigotes derived from Vero cell monolayers. Bugs were dissected 25–30 days later, *T. cruzi* was re-isolated by culture on biphasic blood agar, and was passaged into axenic culture to obtain sufficient organisms for testing drug sensitivities. Last, groups of three immunocompromised (SCID) mice were inoculated with a mixture of faeces from triatomine bugs carrying P1-*hyg* or P2-*neo*. Populations were subsequently retrieved from infected mice into axenic culture and placed under drug pressure.

Determination of phenotype and genotype

DNA purification was carried out by phenol/chloroform extraction and ethanol precipitation or using DNeasy (Qiagen). Amplification reactions used the following conditions: denaturation for 5 min at 94 °C, then 30 cycles of 94 °C (1 min), 50–62 °C (1 min; depending on the primer T_m) and 72 °C (1 min per 1,000 bp), followed by 10 min at 72 °C. Primers are described in Supplementary appendix 6. PGM phenotype determination and RAPD analysis were as described previously⁶. Episomes were detected by multiplex amplification using primers designed to the hygromycin phosphotransferase and neomycin phosphotransferase genes. For karyotype analysis we used a Bio-Rad CHEF Mapper with an autocalgorithm for separation of 0.4–2.2-Mb fragments, followed by Southern blotting and hybridization with radiolabelled probes.

The following DNA sequences were amplified, some with *Taq* Extender; Stratagene: (1) *tpm1*¹⁹ (trypanodoxin: GenBank accession number AF106855; 435 bp); (2) *gpi* (1,038 bp); (3) putative *pgm* (380 bp) (TIGR database (<http://www.tigr.org>) accession number TCI375), identified by similarity between human and putative *Leishmania* *pgm*; (4) *tcp*, an intergenic region (760 bp, including gaps) flanked by 3' *tcp17* and partial 5' *tcpgp2* (ref. 20); (5) a mitochondrial locus (1,078 bp) spanning the maxicircle-encoded genes cytochrome oxidase subunit II (*colII*) and *nd1* (ref. 3); and (6) *dhfr-ts* (1,042 bp)³. PCR products from all loci were cloned into pGEM T (Promega), except where no heterozygous alleles were detectable (mitochondrial DNA³ of progeny; *gpi* of some isolates). For each isolate a minimum of three (*gpi*) and either three or six (*tcp*) clones were sequenced on a capillary sequencer (Beckman) or an ABI 377 using relevant kits.

Genotypes were also determined at eight microsatellite loci¹³, with primers labelled with FAM, NED and HEX, and sized against the ROX 500 marker (ABI) using an ABI 377. An additional 12 microsatellite loci were identified by searching the *T. cruzi* genome database for dinucleotide repeats (TIGR database; see also Supplementary appendix 7). Genescan and Genotyper software (Applied Biosystems) were used to automate measurements of allele length. All microsatellite loci were amplified from P1-*hyg* and P2-*neo* and their double-drug-resistant progeny using standard conditions; a subset (MCLE01, MCLEF10, MCLEG10, SCLF10 and SCLF11 (ref. 13), and A427, A831.3, E801, J356 and N060 (Supplementary appendix 7)) was amplified from all reference strains and field isolates (Supplementary appendix 8).

Phylogenetic analysis

Nucleotide sequences were aligned using Clustal X²⁷ then edited by hand, and are available on request. Primarily, sites of recombination were examined by bootscan analysis²⁸ using the Kimura two-parameter model. Recombination sites were investigated further by maximum likelihood breakpoint analysis²⁹. All maximum likelihood parameter estimates comprising a four-category gamma distribution and a transition/transversion ratio (κ) were obtained with the tree bisection reconnection (TBR) heuristic search (where reconnection limit = 4) using PAUP* 4.0b (D. L. Swofford). Each data set of field isolates and reference strains was also subject to refined split decomposition analysis (Kimura three-parameter model)³⁰.

Received 11 October 2002; accepted 14 January 2003; doi:10.1038/nature01438.

1. Cook, G. C. & Zumla, A. (eds) *Manson's Tropical Diseases* (Saunders, London, 2003).
2. World Health Organisation. *Control of Chagas Disease* (World Health Organisation Technical Report Series 905, Geneva, 2002).
3. Machado, C. A. & Ayala, F. J. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proc. Natl Acad. Sci. USA* **98**, 7396–7401 (2001).
4. Gibson, W. C. & Stevens, J. R. Genetic exchange in the trypanosomatidae. *Adv. Parasitol.* **43**, 1–46 (1999).
5. Bingle, L. E., Eastlake, J. L., Bailey, M. & Gibson, W. C. A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the *in situ* detection of mating events. *Microbiology* **147**, 3231–3240 (2001).
6. Carrasco, H. J., Frame, I. A., Valente, S. A. & Miles, M. A. Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi*. *Am. J. Trop. Med. Hyg.* **54**, 418–424 (1996).
7. Gibson, W. C. & Bailey, M. Genetic exchange in *Trypanosoma brucei*: evidence for meiosis from analysis of a cross between drug-resistant transformants. *Mol. Biochem. Parasitol.* **64**, 241–252 (1996).
8. Stothard, J. R., Frame, I. A. & Miles, M. A. Genetic diversity and genetic exchange in *Trypanosoma cruzi*: dual drug-resistant 'progeny' from episomal transformants. *Mem. Inst. Oswaldo Cruz* **94** Suppl. 1, 189–193 (1999).
9. Ohno, S. *Evolution by Gene Duplication* (Springer, Berlin, 1970).
10. Knight, J. All genomes great and small. *Nature* **417**, 374–376 (2002).
11. Brisse, S., Barnabe, C. & Tibayrenc, M. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *Int. J. Parasitol.* **30**, 35–44 (2000).

12. Mendonça, M. B. *et al.* Two main clusters within *Trypanosoma cruzi* zymodeme 3 are defined by distinct regions of the ribosomal RNA cistron. *Parasitology* **124**, 177–184 (2002).
13. Oliveira, R. P. *et al.* Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. *Proc. Natl Acad. Sci. USA* **95**, 3776–3780 (1998).
14. Miles, M. A. *et al.* Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas disease? *Lancet* **1**, 1338–1340 (1981).
15. Gaunt, M. W. & Miles, M. A. The ecotopes and evolution of triatomine bugs (Triatominae) and their associated trypanosomes. *Mem. Inst. Oswaldo Cruz* **95**, 557–565 (2000).
16. Tibayrenc, M. & Ayala, F. J. The clonal theory of parasitic protozoa: 12 years on. *Trends Parasitol.* **18**, 405–410 (2002).
17. McDaniel, J. P. & Dvorak, J. A. Identification, isolation, and characterization of naturally-occurring *Trypanosoma cruzi* variants. *Mol. Biochem. Parasitol.* **57**, 213–222 (1993).
18. Kelly, J. M. Genetic transformation of parasitic protozoa. *Adv. Parasitol.* **39**, 227–270 (1997).
19. Wilkinson, S. R. *et al.* The *Trypanosoma cruzi* enzyme TcGPXI is a glycosomal peroxidase and can be linked to trypanothione reduction by glutathione or trypanedoxin. *J. Biol. Chem.* **277**, 17062–17071 (2002).
20. Robello, C., Gamarro, F., Castanys, S. & Alvarez-Valin, F. Evolutionary relationships in *Trypanosoma cruzi*: molecular phylogenetics supports the existence of a new major lineage of strains. *Gene* **246**, 331–338 (2000).
21. MacLeod, A. *et al.* Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proc. Natl Acad. Sci. USA* **97**, 13442–13447 (2000).
22. Spratt, B. G. & Maiden, M. C. J. Bacterial population genetics, evolution and epidemiology. *Phil. Trans. R. Soc. Lond. B* **354**, 701–710 (1999).
23. Channanpant, J., Shan, W. X. & Tyler, B. M. High frequency mitotic gene conversion in genetic hybrids of the oomycete *Phytophthora sojae*. *Proc. Natl Acad. Sci. USA* **98**, 14530–14535 (2001).
24. Cruz, A. K., Titus, R. & Beverley, S. M. Plasticity in chromosome number and testing of essential genes in *Leishmania* by targeting. *Proc. Natl Acad. Sci. USA* **90**, 1599–1603 (1993).
25. Gaunt, M. W. & Miles, M. A. A molecular clock for the insects dates the origin of the insects and accords with paleontological and biogeographic landmarks. *Mol. Biol. Evol.* **19**, 748–761 (2002).
26. Miles, M. A. in *Protocols in Molecular Parasitology* (ed. Hyde, J. E.) 15–28 (Humana, Totowa, New Jersey, 1992).
27. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. CLUSTAL _X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
28. Salminen, M. O., Carr, J. K., Burke, D. S. & McCutchan, F. E. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retro.* **11**, 1423–1425 (1995).
29. Holmes, E. C., Worobey, M. & Rambaut, A. Phylogenetic evidence for recombination in dengue virus. *Mol. Biol. Evol.* **16**, 405–409 (1999).
30. Dopazo, J., Dress, A. & Vonhaeseler, A. Split decomposition—a technique to analyse viral evolution. *Proc. Natl Acad. Sci. USA* **90**, 10320–10324 (1993).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements We thank the Wellcome Trust for financial support, D. Conway for valuable advice, and S. Wilkinson, S. Obado and J. Kelly for gifts of primers and comments on the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.A.M. (e-mail: michael.miles@lshtm.ac.uk).

Water transport in plants obeys Murray's law

Katherine A. McCulloh, John S. Sperry & Frederick R. Adler

Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

The optimal water transport system in plants should maximize hydraulic conductance (which is proportional to photosynthesis^{1–5}) for a given investment in transport tissue. To investigate how this optimum may be achieved, we have performed computer simulations of the hydraulic conductance of a branched transport system. Here we show that the optimum network is not achieved by the commonly assumed pipe model of plant form^{6–8}, or its antecedent, da Vinci's rule^{9,10}. In these representations, the number and area of xylem conduits is constant at

every branch rank. Instead, the optimum network has a minimum number of wide conduits at the base that feed an increasing number of narrower conduits distally. This follows from the application of Murray's law, which predicts the optimal taper of blood vessels in the cardiovascular system¹¹. Our measurements of plant xylem indicate that these conduits conform to the Murray's law optimum as long as they do not function additionally as supports for the plant body.

The idea that organisms may be optimally designed has fuelled considerable research and controversy^{12–15}. Both animals and plants require extensive and expensive transport systems, for which an optimal design presumably provides a selective advantage. In plants, transport through xylem has allowed growth in height and colonization of diverse habitats¹⁶, and must be extensive because photosynthesis entails an exchange of large quantities of internal water for atmospheric carbon dioxide. Plants use capillary suction to transport water from the roots to the leaves without using metabolic energy, and the xylem conduits themselves are cheap to maintain, being dead at maturity. However, the numerous thick-walled conduits require considerable carbon investment. This cost, combined with the benefit to photosynthesis of optimizing xylem design, implies that it would be advantageous for plants to maximize hydraulic conductance per vascular investment, and hence, to obey Murray's law¹¹.

In 1926, Murray¹¹ developed a theory of optimal cardiovascular design that solves for the sizes of blood vessels from the aorta through progressive branch points to the capillaries that maximize the hydraulic conductance of flow through the vascular network for a fixed investment in blood and vessel volume. Under ideal conditions, the optimal design equalizes the sum of all radii cubed (Σr^3) at all points along the flow path if the volume flow rate (Q) of the blood is conserved within the vascular system and the flow is laminar¹¹. This result, known as Murray's law, fits data better than equalizing the sum of radii raised to the second or fourth powers^{17–19}. Surprisingly, this law has not been systematically applied and tested in plant vasculature²⁰.

Murray's law is applicable to plant xylem given the following four conditions. (1) The steady-state xylem Q is constant along the flow path. With the exception of the absorbing roots and the minor leaf veins, there will be no net loss of water from the xylem to surrounding tissues under steady-state conditions. (2) Xylem hydraulic conductance is proportional to conduit radius raised to the fourth power. This is the Hagen–Poiseuille prediction for laminar flow through cylindrical capillaries, which are appropriate

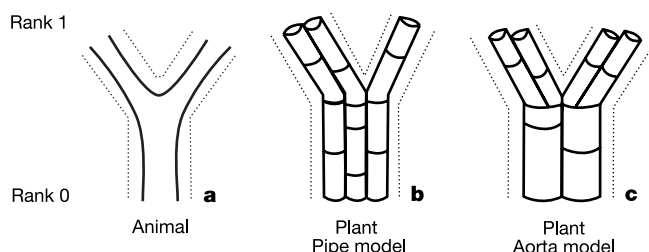


Figure 1 Transport networks. **a**, Vascular networks in animals; **b**, **c**, plant xylem. The most proximal rank (rank 0) branches to form progressively distal ranks (for example, rank 1). In the animal cardiovascular system (**a**), a single continuous tube ramifies and the number of branches increases by an integer factor (the conduit furcation number, F) between ranks. In the plant system, tubes are not continuous, but divided into conduits. Hundreds of conduits run in parallel within each branch (only a few are shown for clarity). In the 'pipe model' (**b**), the number of conduits is identical regardless of rank, and $F = 1$. In the 'aorta model' (**c**) the number of conduits increases with the number of branches by an integer factor as in the animal system, and $F = 2$. The pipe and aorta models are points in a continuum, because with hundreds of conduits in a single branch, F can be a fraction.

approximations for low-velocity xylem flow. (3) Conduit wall volume is proportional to total conduit volume. Murray assumed that the major volume cost was in the blood, and that the thin walls of blood vessels were negligible. In plants it is the opposite: the xylem water is cheap, but the conduit walls are costly. Murray's law is still valid, as long as the conduit wall volume scales directly with total conduit volume. The wall volume/conduit volume is proportional to the cavitation resistance of xylem²¹, as is necessary to withstand the compressive forces on the wall caused by negative pressure. Thus, for a given cavitation resistance throughout the xylem network, Murray's law should apply. (4) The xylem conduits must function primarily in transport, as opposed to providing an additional mechanical support role. When conduits are serving both functions (as in conifer shoots or diffuse-porous trees), the optimization criteria must include mechanical as well as hydraulic considerations, and Murray's law is inappropriate.

Murray's law by itself is not sufficient to characterize the conductance versus investment trade-off in xylem, because of fundamental structural differences between the vasculature of animals and plants. Figure 1a illustrates the single-branched tube of the animal network. As the single parent tube at rank 0 (for example, the

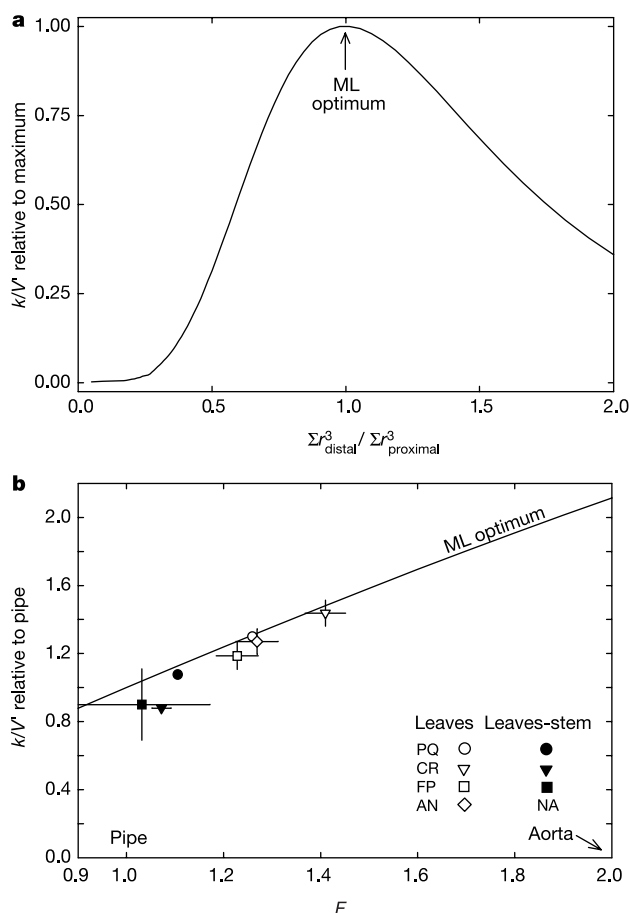


Figure 2 Murray's law. **a**, The Murray's law (ML) optimum. The line shows relative hydraulic conductance (k) per fixed volume (V') plotted against the ratio of the sum of conduit radii cubed (Σr^3) of distal:proximal ranks in a modelled vascular network. The maximum k/V' occurs at a Σr^3 ratio of 1 according to Murray's law regardless of the conduit furcation number (F). **b**, The Murray law optimum versus F . The optimum is set to 1 for $F = 1$ (pipe model). The greater efficiency of wide, proximal conduits in systems with higher F increases the hydraulic conductance for a given volume over the pipe model. The symbols show the measured F relative to the mean Murray's law optimum within leaves (open symbols, $\pm$ s.d.) and comparing petioles with stems (filled symbols, $\pm$ s.d.). PQ, *Parthenocissus quinquefolia* (vine); CR, *Campsis radicans* (vine); FP, *Fraxinus pennsylvanica* (ring-porous tree); and AN, *Acer negundo* (diffuse-porous tree).

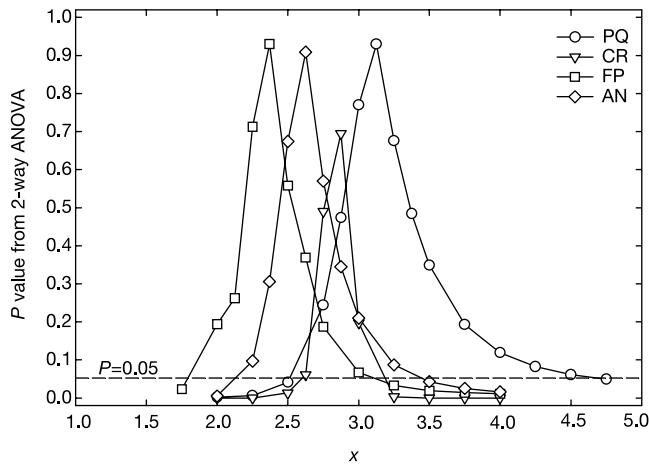


Figure 3 ANOVA P values comparing the sum of the conduit radii raised to the x power (Σr^x) between petiolule versus petiole ranks of compound leaves. The P values are shown versus the exponent x . As the P value approaches 1, differences in Σr^x between ranks are minimized. Although Murray's law predicts the maximum P value at $x = 3$, the law cannot be rejected if P at $x = 3$ is above 0.05 (dashed line). None of the P values at $x = 3$ were below 0.05.

aorta) branches into daughter tubes at rank 1, the number of tubes increases by a predictable integer factor, here referred to as the conduit furcation number (F). Such a system would be disastrous in plants because the xylem sap is under negative pressure, and a single injury would fill the entire network with air. Instead, the main stem of a plant at rank 0 has numerous conduits arranged in parallel and series, which can confine any dysfunction to a single conduit. Unlike the animal system, there is no constraint on F . For example, in Fig. 1b, the number of conduits stays the same across ranks; this configuration has become known as the "pipe model"^{6,22} ($F = 1$). Alternatively, the conduits can increase with branch rank by the same integer factor as in the animal system (Fig. 1c, 'aorta model'), but with multiple 'aortas' at rank 0 ($F = 2$). The pipe versus aorta distinctions are points in a continuum because F need not be an integer.

How does the pipe versus aorta configuration influence the trade-off between vascular conductance and investment? Regardless of F , the modelled maximum hydraulic conductance per vascular volume always occurs when Σr^3 is conserved throughout the system (Fig. 2a). This can also be shown by incorporating F in the derivation of Murray's law (Supplementary Information). However, the absolute value of the optimum increases with F (Fig. 2b). The pipe model yields substantially lower hydraulic conductance than the aorta model. The aorta model achieves greater conductance per volume by exploiting the increased efficiency of fewer, wider conduits in the proximal ranks.

The data in Fig. 2 raise the question of where real plants fall with respect to optimal conductance per volume. We put forward the hypothesis that xylem networks that meet the four assumptions of Murray's law should fall on the optimum with an $F > 1$ to exploit

the efficiency of fewer, wider conduits proximally. This hypothesis was tested in three contexts where the xylem conduit network does not perform a major mechanical support function: compound leaves, vines, and ring-porous trees (Table 1). Leaves are supported by turgor pressure and collenchyma, not xylem conduits. Compound leaves were chosen because there is no change in Q between petiolule and petiole ranks, unlike between ranks within leaf blade venation where Q is lost to transpiration. Vines are structural parasites, being supported by other plants or structures. In ring-porous trees, fibres provide the bulk of the structural support and the vessels comprise a relatively small fraction of the wood²¹.

The data for compound leaves were consistent with Murray's law. Figure 3 shows the results of the petiolule versus petiole comparison in four species (Table 1). The Σr^x comparisons across ranks yielded maximum values of P for x between 2.4 and 3.2. In no species was the P value for $x = 3$ less than 0.05 (Fig. 3, dashed line), meaning that the Σr^3 was not significantly different between ranks. The shoot xylem of vine and ring-porous tree species was also generally consistent with Murray's law. Two out of the three petiolule versus stem comparisons had a P value for $x = 3$ greater than 0.05. The one deviant was only slightly different with the closest x at $P > 0.05$ equal to 2.85 (data not shown).

The proximity of each comparison to the Murray optimum is shown by the symbols in Fig. 2b. All comparisons had $F > 1$ as predicted, and there was a significant trend for xylem closest to Murray's law to have the greatest F value (Kendall's coefficient of rank correlation, $P < 0.05$ (ref. 23), symbols in Fig. 2b). The greatest F of 1.45 fell short of the aorta model ($F = 2$), perhaps because the increase in efficiency is offset by the reduction of conduit number in proximal stems and their increasing diameter (Table 1)—both factors can render the system vulnerable to failure²⁴. Importantly, the situations in which F was greatest always occurred in the leaves, which are constructed specifically as disposable organs. Leaf versus stem comparisons tended to have greater Σr^3 in proximal versus distal ranks, and lower F (Fig. 2b, filled symbols).

Despite considerable interest in the adaptive significance of plant form, we are still far from understanding it. Quantifying the applicability of Murray's law is just one piece of the puzzle. Nevertheless, it improves upon Leonardo da Vinci's 'rule' that the total cross-sectional area of a tree's stems must be constant across branch ranks (area-preserving branching) to adequately supply the leaves^{9,10}. The Murray optimum at $F > 1$ requires an increase in conduit cross-sectional area along the branched flow path (Table 1). The interaction between hydraulic supply and mechanical support of plant canopies undoubtedly underlies the distribution of supply versus support tissue in shoot structure. The two requirements are in conflict, because the optimal distribution of a fixed volume of tissue to support the tallest column requires a reduction in tissue area with height^{25,26}.

The allometry of plant form predicts large-scale biological consequences—from the scaling between physiological processes and body size to the energy budgets of ecosystems. Recent work shows how a very specific plant vascular organization (the pipe model with $F = 1$) and branching pattern (exact self-similarity with no apical dominance) that functions at the equivalent of the Murray's law

Table 1 Conduit number, mean radii and area

Rank	PQ (vine, $n = 3$)			CR (vine, $n = 3$)			FP (ring-porous, $n = 3$)			AN (diffuse-porous, $n = 4$)		
	CN	r (μm)	CA	CN	r (μm)	CA	CN	r (μm)	CA	CN	r (μm)	CA
Petiolule	10.4 (0.4)	17.08 (0.3)	2.6 (0.1)	17.7 (3.0)	13 (1.0)	2.5 (0.7)	3.0 (1.2)	19.9 (1.2)	1.1 (0.3)	1.8 (0.1)	18.4 (0.9)	1.2 (0.1)
Petiole	5.2 (0.1)	21.7 (0.6)	2.1 (0.2)	3.5 (0.6)	21.7 (1.8)	1.4 (0.3)	1.5 (0.6)	26.7 (2.1)	1 (0.3)	1 (0)	23.8 (2.3)	1 (0)
1-yr bud scar	1 (0)	35.8 (1.9)	1 (0)	1 (0)	32.6 (5.6)	1 (0)	1 (0)	24 (5)	1 (0)	NA	NA	NA

Branch ranks, relative change in total conduit number versus most proximal rank (CN), average conduit radius (r), and relative change in total conduit area versus most proximal rank (CA) for each species. The standard deviation from the mean of the individuals is listed in parentheses. PQ, *Parthenocissus quinquefolia* (vine); CR, *Campsis radicans* (vine); FP, *Fraxinus pensylvanica* (ring-porous tree); AN, *Acer negundo* (diffuse-porous tree). n , number of individuals measured per species. NA, not applicable. The 1-yr-old wood of *A. negundo* was not measured because the wood area is largely composed of conduits, which should contribute to the structural support of the plant.

optimum can be extrapolated to explain metabolic scaling rules in biology^{6–8}. The challenge is to relax the unrealistic structural assumptions of these models and evaluate hydraulic and mechanical optima. We have seen that F is not necessarily 1, and we have determined the applicability of Murray's law to xylem. A challenge for the future is to develop a better understanding of the interaction between tree biomechanics, hydraulics and growth. □

Methods

A computer programme calculated the hydraulic conductance per fixed vascular volume of a branched transport network. We modelled a constant-volume vascular network with specified number of ranks, a branch furcation number, B ($B = 2$, dichotomous, $B = 3$, trichotomous, and so on), and F . We standardized F to equal 1 for the pipe model and 2 for the aorta model regardless of B ($F = [(F' - 1)/(B - 1)] + 1$, where F' is the raw conduit furcation number). The network had constant F across ranks, and constant fractional taper in conduit diameter across ranks (diameter was constant within rank). The number of conduits at the distal rank was one per branch—the minimum required to vascularize the system. Rank length was unity, because the Murray's law solution is independent of branch length. To generate Fig. 2a, we varied conduit taper and computed the network conductance. To generate Fig. 2b, conduit taper was set to match Murray's law, and network hydraulic conductance computed as F was changed.

Conduit measurements were made on species with compound leaves (Table 1) collected in the greater Salt Lake City (Utah) area (40° 46' N, 111° 58' W). Three to four individuals per species were collected from the same habitat, and were similar in size. Upon collection, plants were perfused with 0.05% basic fuchsin to determine which xylem conduits were functional, and only these were measured. Between 2,000 and 6,000 conduits were measured per individual. Conduit radii were measured at the petiole and petiole rank for leaves, and at the current year's annual bud scar in stems (Table 1).

Conduit statistics per rank had to be estimated, given the large numbers of conduits involved. In leaves, all conduits in petioles and petiolules were measured in a sample of ≥ 5 leaves. Each measured leaf provided a Σr^x per leaf area which was multiplied by total leaf area to estimate rank Σr^x . In secondary stem xylem, conduits in ≥ 3 radial sectors per branch were measured, with each sector yielding a Σr^x per sector area. Each subsample was multiplied by total xylem area of the rank to obtain a rank Σr^x estimate. These multiple Σr^x estimates per rank were incorporated into the analysis of variance (ANOVA).

To test Murray's law, the Σr^x of the distal-most rank (petiolule) was compared against the Σr^x of each proximal rank (petiole, stem xylem at first year bud scar). To test the conservation of Σr^3 , x in each of the rank estimates of Σr^x was incremented from 1 to 5. Raising the Σr^x to relatively high x resulted in frequency distributions that were not normally distributed, so the data were log-transformed for statistical analyses. For each species we used a 2-way ANOVA with the Σr^x estimates as the dependent variable, the distal-most versus proximal rank pair as the fixed factor, and the replicate individuals as a random factor (for the petiolule: petiole comparison there were 26 degrees of freedom for PQ, CR and FP; and 45 for AN (see Table 1 for nomenclature); for the petiolule:1-yr-old wood comparison, there were 25 degrees of freedom for PQ, 21 for CR and 43 for FP²⁷ (SPSS Version 8.0.0 (SPSS, Chicago, 1998)). Increased similarity in estimates of Σr^x between ranks resulted in P -values approaching 1. For each paired comparison we used a 2-way ANOVA to identify the range of x over which Σr^x between ranks was not significantly different at $P \geq 0.05$. If this range of x did not include 3, Murray's law was rejected for that comparison.

Received 9 September 2002; accepted 21 January 2003; doi:10.1038/nature01444.

- Hubbard, R. M., Stiller, V., Ryan, M. G. & Sperry, J. S. Stomatal conductance and photosynthesis vary linearly with plant hydraulic conductance in ponderosa pine. *Plant Cell Environ.* **24**, 113–121 (2001).
- Saliendra, N. Z., Sperry, J. S. & Comstock, J. P. Influence of leaf water status on stomatal response to humidity, hydraulic conductance, and soil drought in *Betula occidentalis*. *Planta* **196**, 357–366 (1995).
- Meinzer, F. C. *et al.* Environmental and physiological regulation of transpiration in tropical forest gap species: The influence of boundary layer and hydraulic properties. *Oecologia* **101**, 514–522 (1995).
- Sperry, J. S., Alder, N. N. & Eastlack, S. E. The effect of reduced hydraulic conductance on stomatal conductance and xylem cavitation. *J. Exp. Bot.* **44**, 1075–1082 (1993).
- Meinzer, F. C. & Grantz, D. A. Stomatal and hydraulic conductance in growing sugarcane: Stomatal adjustment to water transport capacity. *Plant Cell Environ.* **13**, 383–388 (1990).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for the structure and allometry of plant vascular systems. *Nature* **400**, 664–667 (1999).
- Enquist, B. J., West, G. B. & Brown, J. H. in *Scaling in Biology* (eds Brown, J. H. & West, G. B.) 167–198 (Oxford Univ. Press, Oxford, 2000).
- Richter, J. P. *The Notebooks of Leonardo da Vinci (1452–1519), Compiled and Edited from the Original Manuscripts* (Dover, New York, 1970).
- Horn, H. S. in *Scaling in Biology* (eds Brown, J. H. & West, G. B.) 199–220 (Oxford Univ. Press, Oxford, 2000).
- Murray, C. D. The physiological principle of minimum work. I. The vascular system and the cost of blood volume. *Proc. Natl Acad. Sci. USA* **12**, 207–214 (1926).
- Horn, H. S. *The Adaptive Geometry of Trees* (Princeton Univ. Press, Princeton, New Jersey, 1971).
- Givnish, T. J. in *On the Economy of Plant Form and F* (ed. Givnish, T. J.) 3–9 (Cambridge Univ. Press, Cambridge, 1986).
- Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598 (1979).
- Mark, R. Architecture and evolution. *Am. Sci.* **84**, 383–389 (1996).

- Raven, J. A. The evolution of vascular land plants in relation to supracellular transport processes. *Adv. Bot. Res.* **5**, 153–219 (1987).
- Sherman, T. F. On connecting large vessels to small: The meaning of Murray's law. *J. Gen. Physiol.* **78**, 431–453 (1981).
- LaBarbera, M. Principles of design of fluid transport systems in zoology. *Science* **249**, 992–999 (1990).
- Vogel, S. *Life in Moving Fluids: The Physical Biology of Flow* (Princeton Univ. Press, Princeton, New Jersey, 1994).
- Canny, M. J. The transpiration stream in the leaf apoplast: Water and solutes. *Phil. Trans. R. Soc. Lond. B* **341**, 87–100 (1993).
- Hacke, U. G., Sperry, J. S., Pockman, W. P., Davis, S. D. & McCulloh, K. A. Trends in wood density and structure are linked to prevention of xylem implosion by negative pressure. *Oecologia* **126**, 457–461 (2001).
- Shinozaki, K., Yoda, K., Hozumi, K. & Kira, T. A quantitative analysis of plant form—the pipe model theory: I. Basic analysis. *Jpn. J. Ecol.* **14**, 97–105 (1964).
- Sokal, R. R. & Rohlf, F. J. *Biometry: The Principles and Practice of Statistics in Biological Research* (Freeman, New York, 1995).
- Zimmermann, M. H. in *Xylem Structure and the Ascent of Sap* (ed. Timell, T. E.) 15–16, 99 (Springer, Berlin, 1983).
- Keller, J. B. & Niordson, F. I. The tallest column. *J. Math. Mech.* **16**, 433–446 (1966).
- McMahon, T. A. Size and shape in biology. *Science* **179**, 1201–1204 (1973).
- Grafen, A. & Hails, R. *Modern Statistics for the Life Sciences* (Oxford Univ. Press, Oxford, 2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank A. Collopy and M. McCord for assistance in collecting plants. This work was partly supported by Sigma Xi (K.A.M.) and NSF (J.S.S.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.A.M. (e-mail: mcculloh@biology.utah.edu).

Extra-embryonic function of Rb is essential for embryonic development and viability

Lizhao Wu*, Alain de Bruin*, Harold I. Saavedra*, Maja Starovic†, Anthony Trimboli*, Ying Yang‡, Jana Opavska*, Pamela Wilson§, John C. Thompson§, Michael C. Ostrowski||, Thomas J. Rosol||, Laura A. Woollett#, Michael Weinstein§||, James C. Cross†, Michael L. Robinson‡|| ☆ & Gustavo Leone*§||

* Human Cancer Genetics Program, Department of Molecular Virology, Immunology and Medical Genetics, § Department of Molecular Genetics, || Comprehensive Cancer Center, ¶ Department of Veterinary Biosciences, and ☆ Department of Pediatrics, The Ohio State University, Columbus, Ohio 43210, USA
† Department of Biochemistry & Molecular Biology, University of Calgary Faculty of Medicine, Calgary, Alberta T2N 4N1, Canada
‡ Division of Molecular and Human Genetics, Children's Research Institute, Columbus, Ohio 43205, USA
Department of Pathology and Laboratory Medicine, University of Cincinnati Medical Center, Cincinnati, Ohio 45267, USA

The retinoblastoma (*Rb*) gene was the first tumour suppressor identified¹. Inactivation of *Rb* in mice results in unscheduled cell proliferation, apoptosis and widespread developmental defects, leading to embryonic death by day 14.5 (refs 2–4). However, the actual cause of the embryonic lethality has not been fully investigated. Here we show that loss of *Rb* leads to excessive proliferation of trophoblast cells and a severe disruption of the normal labyrinth architecture in the placenta. This is accompanied by a decrease in vascularization and a reduction in placental transport function. We used two complementary techniques—tetraploid aggregation and conditional knockout strategies—to demonstrate that *Rb*-deficient embryos supplied

with a wild-type placenta can be carried to term, but die soon after birth. Most of the neurological and erythroid abnormalities thought to be responsible for the embryonic lethality of *Rb*-null animals were virtually absent in rescued *Rb*-null pups. These findings identify and define a key function of *Rb* in extra-embryonic cell lineages that is required for embryonic development and viability, and provide a mechanism for the cell autonomous versus non-cell autonomous roles of *Rb* in development.

During mammalian development, one of the earliest differentiation events is the decision to commit to either an extra-embryonic cell lineage or to the inner-cell mass that eventually makes up the embryo proper. Previous studies suggest that p57^{Kip2} and p27^{Kip1}, two principal cell-cycle regulators in the *Rb* pathway, have an important role in the control of proliferation and differentiation of extra-embryonic trophoblast cells within the labyrinth layer of the placenta⁵. More recently we have found that ablation of E2F3, a downstream effector of *Rb* function, also leads to cellular abnormalities in the placenta (A.d.B. and G.L., unpublished observations). Given that loss of either upstream regulators (p57^{Kip2} and p27^{Kip1}) or downstream effectors (E2F3) of *Rb* function leads to cellular proliferation defects in the placenta, we hypothesized that disruption of *Rb* itself might also have deleterious consequences in placental development.

To test this hypothesis, we first analysed haematoxylin-and-eosin-stained sections obtained from wild-type and *Rb*-deficient placentae. At embryonic day (E)13.5, a time when *Rb* mutant embryos begin to die, the three layers of the placenta—labyrinth, spongiotrophoblast and trophoblast giant cell layers—are easily discernible (Fig. 1a, and data not shown). The labyrinth is the site of oxygen and nutrient exchange between the mother and fetus, and consists of highly branched, trophoblast-covered maternal blood sinusoids in direct contact with the underlying fetal blood vessels. The wild-type labyrinth therefore has a porous appearance, with trophoblast cells being well organized and distributed around fetal and maternal blood vessels (Fig. 1b, *Rb*^{+/+}). The *Rb* mutant placenta was of normal size and the trophoblast giant cell and spongiotrophoblast layers appeared normal. However, the labyrinth layer was severely disrupted with an accumulation of large clusters of densely packed trophoblast cells at the expense of maternal and fetal blood spaces (Fig. 1a, b, *Rb*^{-/-}). Similar architectural defects were evident at E12.5 (Supplementary Fig. 1a, b). This abnormal phenotype was completely penetrant, as all of the seventeen E12.5 and E13.5 *Rb*^{-/-} placentae examined, but none of the wild-type controls, exhibited the same cellular defect. At E11.5, however, the differences between *Rb* mutant and wild-type placentae were variable and difficult to discern reliably, probably due to the highly proliferative status of labyrinth trophoblasts in wild-type placentae at this stage (Fig. 2f; see also Supplementary Fig. 1c, d).

To determine whether the defect in *Rb*-deficient placentae was restricted to a specific cell type, we performed marker analyses on 5-μm sections from E13.5 placentae. Two giant-cell-specific markers, *Pl1* (ref. 6) and proliferin⁷, were expressed in the correct spatial pattern (data not shown). In addition, as determined by the spongiotrophoblast-specific expression of *Tpbb*⁸, the spongiotrophoblast and the underlying labyrinth layer in *Rb*^{-/-} placentae were well demarcated (Fig. 1c). However, the trophoblast populations within the labyrinth layer were abnormal. In *Rb*^{+/+} placentae, *Eomes* expression was localized normally in clusters of small cuboidal trophoblasts having the typical morphology of trophoblast stem (TS) cells^{9,10}. In contrast, *Eomes*-expressing cells were more abundant and broadly distributed throughout the labyrinth of *Rb*^{-/-} placentae (Fig. 1d), suggesting an abnormal expansion of TS cells and perhaps a delay in differentiation.

Because there was an obvious decrease in maternal and fetal blood spaces in the labyrinth of *Rb* mutant placentae, a number of histological assays were used to characterize this defect further. In

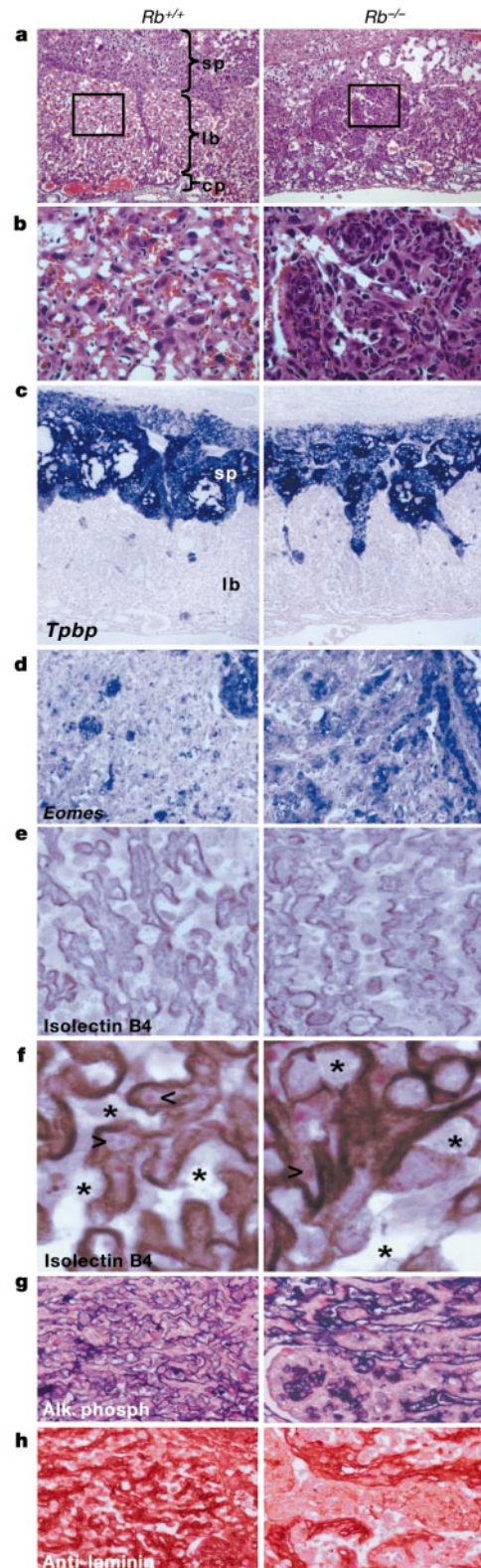


Figure 1 *Rb*^{-/-} mice show placental dysplasia in the labyrinth layer. **a, b**, Haematoxylin and eosin staining of placental sections from E13.5 wild-type (*Rb*^{+/+}) and mutant (*Rb*^{-/-}) animals. The boxed areas in **a** are shown at higher magnification in **b**. The giant cell layer, which is adjacent to the spongiotrophoblast layer, is not shown. **c–h**, Cell lineage marker analyses of E13.5 placental sections by *in situ* hybridization (**c, d**) or histochemistry (**e–h**). Arrows and asterisks denote fetal blood spaces and maternal blood spaces, respectively. cp, chorionic plate; lb, labyrinth; sp, spongiotrophoblast. Original magnification: $\times 10$ (**a, c**), $\times 20$ (**b, d, g, h**), $\times 40$ (**e**), $\times 100$ (**f**).

wild-type placentae, trophoblast cells lining the maternal blood spaces are elongated and thin, and express, at the brush border, endogenous alkaline phosphatase activity as well as a cell-surface carbohydrate structure detected with isolectin B4 (Fig. 1e, f, $Rb^{+/+}$).

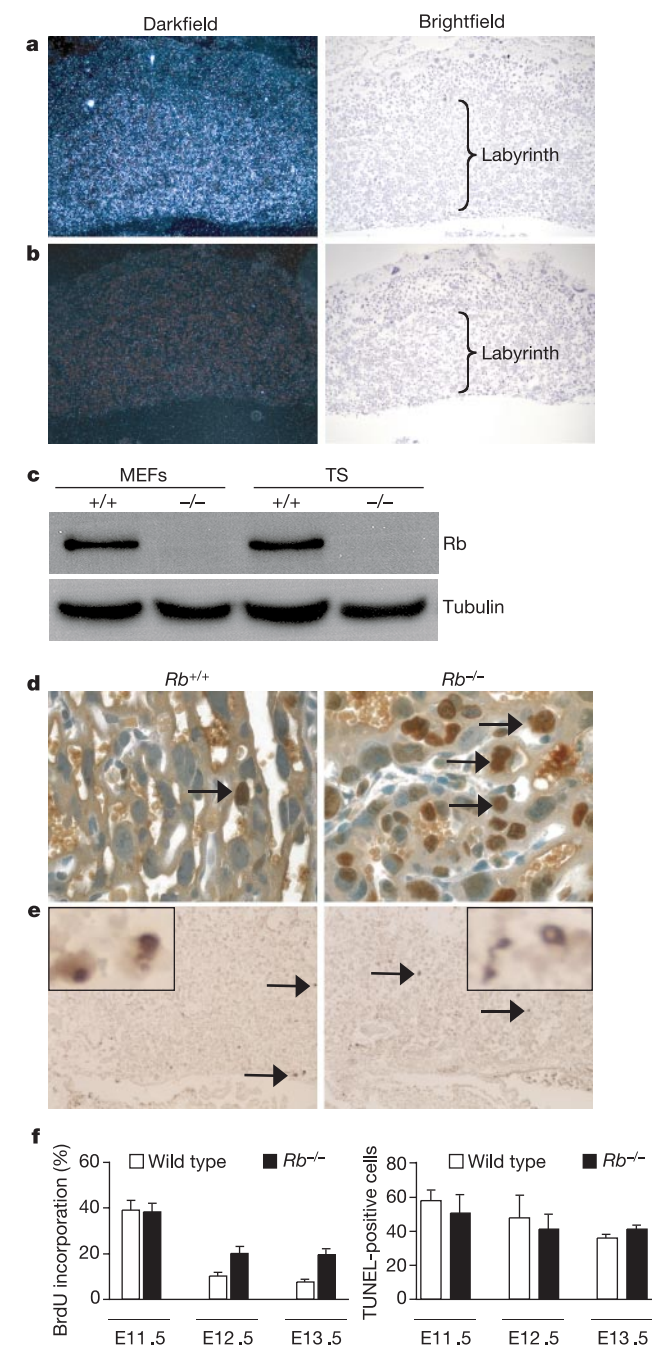


Figure 2 Loss of *Rb* results in higher levels of proliferation in the trophoblast cells. **a, b**, *In situ* hybridization of E12.5 placental sections using *Rb* antisense (**a**) or sense probes (**b**). **c**, Western blot analysis of wild-type (+/+) and *Rb*-deficient (-/-) trophoblast stem (TS) cells and mouse embryonic fibroblasts (MEFs). **d**, Immunohistochemical staining with 5-bromodeoxyuridine (BrdU) on sections from E13.5 wild-type ($Rb^{+/+}$) and *Rb* mutant ($Rb^{-/-}$) placentae. Sections were counter-stained with haematoxylin. **e**, TdT-mediated dUTP nick end-labelling (TUNEL) assays on sections from E13.5 wild-type and *Rb* mutant placentae. Arrows indicate examples of BrdU-positive cells (**d**) and TUNEL-positive cells (**e**). Original magnification: $\times 40$ (**a**), $\times 10$ (**b**), $\times 100$ (inset in **e**). **f**, Quantification of the BrdU immunohistochemical staining (left panel) and TUNEL assays (right panel) on various littermates at different ages. Error bars represent standard deviations.

In *Rb*-deficient placentae fewer of these cells were present, suggesting a failure of trophoblast differentiation within the labyrinth. Morphometric analysis of the alkaline phosphatase-positive cells showed that the trophoblast transport surface area within the labyrinth was reduced by 38%, indicating a much less extensive villous structure (Fig. 1g; $77.12 \pm 1.56\%$ compared with $48.00 \pm 2.57\%$ in wild-type and mutant placentae, respectively; $P < 0.0001$). On the other hand, the density of fetal capillaries, as shown by laminin immunoreactivity, was decreased by 12% relative to wild type (Fig. 1h; $75.25 \pm 1.86\%$ compared with $66.31 \pm 2.70\%$; $P < 0.011$). Given that fetal capillaries are limited to the maternal villous structure created by the branching of the trophoblast epithelium, the relative capillary density is therefore higher than normal. This analysis confirms that *Rb* mutant placentae have decreased vascular capacity, but as the decrease in fetal capillary density is less pronounced than the decrease in the surface area occupied by trophoblasts lining the blood spaces, a primary vascular defect is unlikely.

The over-representation of trophoblast cells and the decrease of blood spaces in the *Rb*-deficient labyrinth indicates restricted nutrient and/or oxygen exchange between the mother and the developing embryos. To assess directly placental transport, we measured the accumulation of essential fatty acids (EFAs) in the fetuses and placentae of *Rb* mutant and wild-type littermate E13.5 embryos from four litters by gas chromatographic analysis¹¹. EFAs are obtained from the diet, and thus the fetal content of EFAs is a direct reflection of placental transport capacity. As shown in Table 1, *Rb*-deficient fetuses had significantly lower amounts of the three EFAs (linoleic acid (18:2), arachidonic acid (20:4) and docosahexaenoic acid (22:6)), but had normal levels of non-essential fatty acids (palmytic acid (16:0) and stearic acid (18:0)). As fatty acid composition of the placenta was not affected by loss of *Rb* (Table 1), the decrease of EFAs in *Rb*^{-/-} fetuses suggests a malfunctioning placenta. Considering that the placenta becomes functional for transport at approximately E11.5 and that prior uptake of EFAs is largely independent of placental function, the reduced levels of EFAs in *Rb* mutant embryos at E13.5 are surprisingly acute.

Consistent with a role for *Rb* in regulating labyrinth development, *in situ* hybridization of E12.5 placental sections revealed the expression of *Rb* messenger RNA in the entire placenta, but with the strongest signal confined to the labyrinth layer (Fig. 2a, b). As marker analysis indicated an abnormal expansion of TS cells within the labyrinth of *Rb* mutant placentae, *Rb* expression was directly assessed in TS cells. To this end, TS cells derived from wild-type or *Rb* mutant E3.5 embryos were isolated and cultured as previously described¹⁰. Western blot analysis of extracts from these cells revealed a significant amount of *Rb* protein in wild-type TS cells that was comparable to that found in mouse embryonic fibroblasts (MEFs) (Fig. 2c). The absence of *Rb* protein in extracts derived from *Rb*^{-/-} TS cells or MEFs verified the identity of *Rb* by this analysis.

Given the well-documented functions of *Rb* in regulating cell-cycle entry and programmed cell death, we measured cellular proliferation and apoptosis in $Rb^{+/+}$ and $Rb^{-/-}$ placentae by 5-bromodeoxyuridine (BrdU) incorporation and TdT-mediated dUTP nick end-labelling (TUNEL) assays, respectively. Analysis of the labyrinth from E13.5 placentae indicated that $Rb^{-/-}$ tropho-

Table 1 Essential fatty acid transport in *Rb*^{-/-} placentae

Fatty acid	Reduction in fetuses (%)	Reduction in placentae (%)
16:0	- 0.8 (>0.05)	2.9 (>0.05)
18:0	2.1 (>0.05)	0.4 (>0.05)
18:2*	15.0 (<0.01)	0 (>0.05)
20:4*	9.8 (<0.001)	0.4 (>0.05)
22:6*	17.5 (<0.001)	-2.4 (>0.05)

See text for full names of the fatty acids.

*Essential fatty acids.

blast cells had a higher BrdU index than wild-type littermate controls (Fig. 2d, f). A similar increase in trophoblast proliferation was observed in $Rb^{-/-}$ placentae as early as E12.5 (Fig. 2f). Notably, small cuboidal trophoblast cells resembling TS cells could be visualized readily by staining adjacent sections with haematoxylin and eosin, and appeared to be preferentially labelled with BrdU (Supplementary Fig. 2). In contrast, no significant differences in the proliferation of spongiotrophoblast cells or trophoblast giant cells were found between $Rb^{-/-}$ and wild-type placentae. Finally, TUNEL assays revealed similar levels of apoptosis in $Rb^{-/-}$ and littermate control placentae at E11.5, E12.5 and E13.5 (Fig. 2e, f), suggesting that the placental abnormalities observed in $Rb^{-/-}$ embryos are probably due to the over-proliferation of TS cells and a failure of them to differentiate.

Considering the severity and timing of the placental phenotype in $Rb^{-/-}$ embryos, we hypothesized that a faulty $Rb^{-/-}$ placenta may be the cause of the embryonic lethality. To explore this possibility, we supplied $Rb^{-/-}$ embryos with normal placentae by two separate but complementary approaches, and analysed the resulting offspring at various developmental stages. In the first approach, wild-

type tetraploid embryos were aggregated with diploid embryos derived from Rb heterozygote intercrosses. In this setting, wild-type tetraploid cells contribute exclusively to the trophoblast cells of the placenta and endoderm of the yolk sac, whereas diploid cells can contribute to the extra-embryonic structures but, importantly, make up the entire embryo proper¹². Control tetraploid aggregation experiments using tetraploid *Rosa26* and diploid wild-type blastocysts demonstrated that β -galactosidase-positive tetraploid cells were present only in extra-embryonic lineages, and were absent from the embryo proper (data not shown). Notably, the wild-type reconstituted placentae associated with either $Rb^{-/-}$, $Rb^{+/-}$ or $Rb^{+/+}$ fetuses appeared normal and had similar low levels of BrdU incorporation in labyrinth trophoblast cells (Fig. 3a, c, and data not shown). Furthermore, semi-quantitative polymerase chain reaction (PCR) analysis comparing genomic DNA from these $Rb^{-/-}$ fetuses with mixed control DNA templates confirmed the exclusion of wild-type tetraploid cells from the embryo proper (less than 0.1% contribution, Fig. 3d). Using this strategy, we recovered several live and apparently normal $Rb^{-/-}$ embryos at E14.5 (Table 2). Several mutant embryos were also recovered at developmental

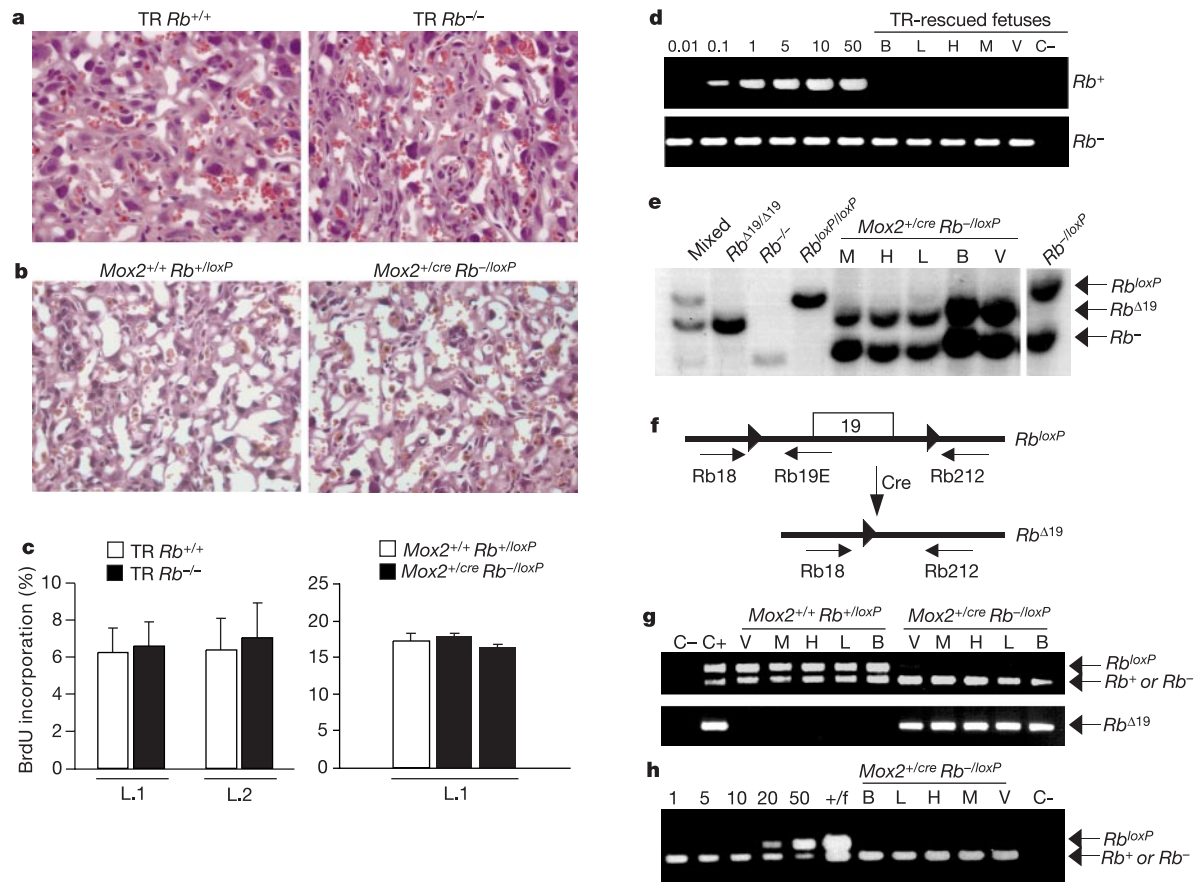


Figure 3 Normal placentae and Rb -deficient fetuses reconstituted by either tetraploid aggregation or conditional knockout approaches. **a**, **b**, Haematoxylin-and-eosin-stained sections from E13.5 reconstituted placentae that were associated with either a 'wild-type' (TR $Rb^{+/+}$ or $Mox2^{+/+} Rb^{+/loxP}$) or an Rb mutant embryo (TR $Rb^{-/-}$ or $Mox2^{+/cre} Rb^{-/loxP}$). Original magnification: $\times 40$. TR, tetraploid. **c**, Quantification of BrdU immunohistochemical-stained sections from E13.5 reconstituted placentae by either tetraploid aggregation (in two litters, L.1 and L.2; left panel) or conditional knockout approaches (in one litter, L.1; right panel). **d**, PCR of genomic DNA from tetraploid rescued embryos as well as mixed DNA templates using specific primer pairs amplifying either the Rb wild-type, Rb^{-} and Rb^{loxP} alleles (top panel) or the $Rb^{\Delta 19}$ allele (bottom panel). Numbers above the lanes represent the percentage of wild-type DNA in the wild-type/ Rb

mutant mixed DNA templates. **e**, Southern blot of various organs from a $Mox2^{+/cre} Rb^{-/loxP}$ newborn mouse. B, brain; H, heart; L, lung; M, muscle; V, liver. **f**, Schematic diagram of the PCR strategy used to detect Cre-mediated deletion of exon 19 from the Rb^{loxP} allele. **g**, PCR of genomic DNA from several organs of newborn mice using primers Rb18 and Rb19E (top panel) or primers Rb18 and Rb212 (bottom panel). C- , water control; C+ , $Rb^{+/loxP}$ (top panel) or $Mox2^{+/cre} Rb^{-/loxP}$ (bottom panel). **h**, PCR of genomic DNA from multiple organs of a $Mox2^{+/cre} Rb^{-/loxP}$ newborn mouse as well as mixed DNA templates using primers Rb18 and Rb19E. Lanes with numbers are $Rb^{loxP/loxP}/Rb^{\Delta 19/\Delta 19}/Rb^{-/-}$ mixed DNA templates with $Rb^{-/-}$ DNA being always 50%, and the numbers above each lane representing the percentage of $Rb^{loxP/loxP}$ DNA in $Rb^{loxP/loxP}/Rb^{\Delta 19/\Delta 19}$ mixtures.

Table 2 Number of embryos recovered by tetraploid aggregation

Age	+/+	+/-	-/-	Total
E14.5	11	10	6 (1)	27
E16.5	3 (1)	5	7 (1)	15
E19.5	5	8 (1)	3 (3)	16
Total	19 (1)	23 (1)	16 (5)	58

Numbers in parentheses represent numbers of dead or reabsorbed embryos recovered.

stages where no $Rb^{-/-}$ embryos arising from natural matings can normally be found²⁻⁴, including seven at E16.5 and three at E19.5, just before birth.

To complement the tetraploid aggregation studies described above, we took advantage of $Rb^{loxP/loxP}$ conditional knockout mice¹³ and $Mox2^{+/cre}$ transgenic mice¹⁴ to reconstitute Rb -deficient embryos with functionally 'normal' placentae. Previous studies have shown that the *Cre* recombinase gene under the control of the endogenous *Mox2* promoter ($Mox2^{+/cre}$) is efficiently expressed and functionally active in all cells of the E6.5 embryo proper, with no expression in the trophoblast or extra-embryonic endoderm lineages¹⁴. β -galactosidase staining of E9.5 and E13.5 embryos and placentae derived from crosses between $Mox2^{+/cre}$ mice and $Rosa26^{loxP}$ reporter mice¹⁵ confirmed the exclusive expression of *Cre* in fetal tissues as previously reported (Supplementary Fig. 3 and data not shown).

By breeding $Mox2^{+/cre}$ mice harbouring one Rb -null allele ($Mox2^{+/cre} Rb^{+/-}$) with homozygous $Rb^{loxP/loxP}$ conditional knockout mice, we were able to generate Rb -deficient embryos ($Mox2^{+/cre} Rb^{-/loxP}$) having placentae of normal histological appearance (Fig. 3b). BrdU incorporation assays revealed no significant differences in DNA replication of trophoblast cells in $Mox2^{+/+} Rb^{+/loxP}$ and $Mox2^{+/cre} Rb^{-/loxP}$ mice (Fig. 3c), suggesting that as in the tetraploid aggregation experiments, the contribution of Rb -null cells to the placenta was minimal. Notably, the Southern blots shown in Fig. 3e demonstrated that although each of the Rb alleles, Rb^{-} , Rb^{loxP} and Rb^{A19} could be readily detected in control and mixed DNA samples, only the Rb^{-} and Rb^{A19} alleles could be detected in the various organs of $Mox2^{+/cre} Rb^{-/loxP}$ newborn pups. Similarly, PCR-based strategies confirmed the *Cre*-mediated deletion of exon 19 from the floxed Rb allele (Rb^{loxP} to Rb^{A19}) in $Mox2^{+/cre} Rb^{-/loxP}$ newborn pups (Fig. 3f, g). Moreover, semi-quantitative PCR analysis using mixed DNA templates as controls demonstrated that in most major organs analysed, the efficiency of *Cre*-mediated deletion of the Rb^{loxP} allele was at least 99%, and in no case was it less than 95% (Fig. 3h). By taking this genetic approach, $Mox2^{+/cre} Rb^{-/loxP}$ fetuses could be recovered at E13.5, E15.5, E18.5 and at birth (Table 3). Notably, $Mox2^{+/cre} Rb^{-/loxP}$ pups were born at near mendelian ratios (25 out of 129 offspring; 19.5% frequency compared to the expected 25% frequency), but died soon after birth. Taken together, tetraploid and genetic rescue experiments suggest that *Rb* has a critical function in extra-embryonic cell lineages that is necessary for fetal viability.

Notably, Rb -deficient animals reconstituted with a functionally normal placenta, either by tetraploid aggregation or by conditional knockout approaches, lack many of the hallmark phenotypes of Rb knockout mice. Histological analysis revealed no apparent defects in the nervous system and haematopoietic compartments.

Table 3 Number of animals recovered by genetic rescue

Age	<i>Mox2</i> ^{+/cre}		<i>Mox2</i> ^{+/+}		Total
	+/ <i>loxP</i>	-/ <i>loxP</i>	+/ <i>loxP</i>	-/ <i>loxP</i>	
E13.5	2	6	4	3	15
E15.5	15	3	6	11	35
E18.5	4	7	8	9	28
Newborn	33	25	40	31	129

In particular, $Mox2^{+/cre} Rb^{-/loxP}$ newborn mice have mostly normal neuronal cells with an increased amount of cytoplasm and few fragmented nuclei (Fig. 4a, c, and data not shown). Moreover, dorsal root ganglia, brains, retinas and spinal cords of these animals appeared histologically normal when compared to those of control counterparts (Fig. 4a–c, and data not shown). The liver phenotype of $Rb^{-/-}$ animals, characterized by increased numbers of nucleated red blood cells and hepatocellular hypoplasia, was almost completely reversed by providing embryos with a functional placenta (Fig. 4d, e). There was, however, a low but significant fraction of erythrocytes that were nucleated in blood smears prepared from $Mox2^{+/cre} Rb^{-/loxP}$ pups compared with that of littermate $Mox2^{+/+} Rb^{+/loxP}$ normal controls (0.49% and 0.08%, respectively; $P < 0.001$; Fig. 4e), suggesting a mild delay in differentiation of the erythroid lineage. In contrast to the lack of obvious defects in

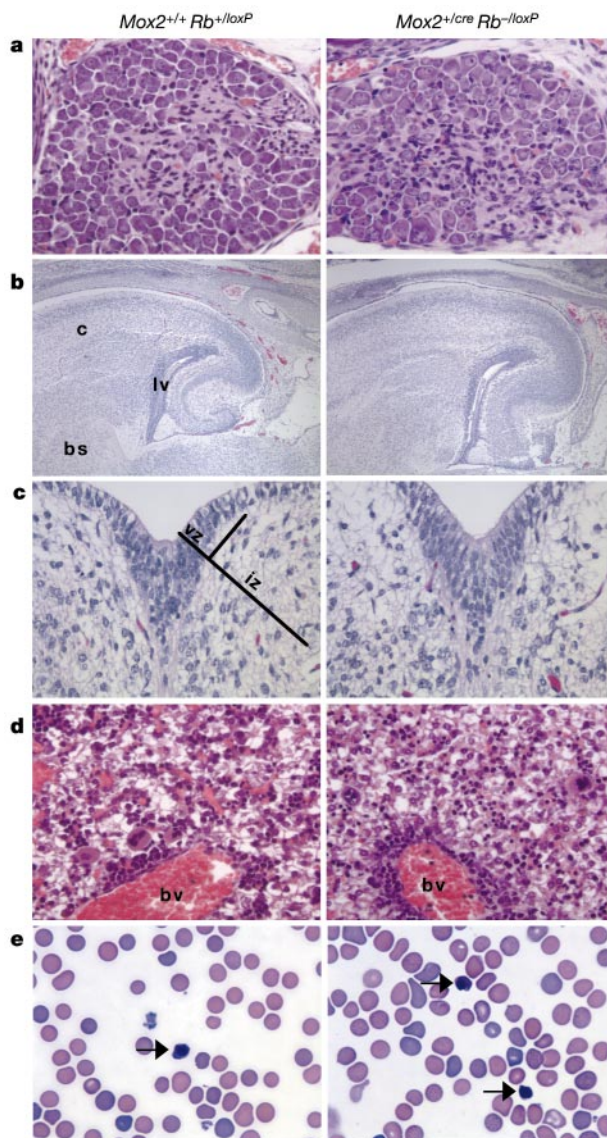


Figure 4 A wild-type placenta suppresses neurogenic and erythropoietic defects of Rb -deficient animals. **a–d**, Histological sections of various tissues from $Mox2^{+/+} Rb^{+/loxP}$ and $Mox2^{+/cre} Rb^{-/loxP}$ newborn mice. **a**, Cervical dorsal root ganglia (sagittal); **b**, midbrain (transverse); **c**, hindbrain (fourth ventricle, transverse); **d**, liver. bs, brainstem; bv, blood vessels; c, cerebrum; iz, intermediate zone; lv, lateral ventricle; vz, ventricular zone. **e**, Blood smear from $Mox2^{+/+} Rb^{+/loxP}$ and $Mox2^{+/cre} Rb^{-/loxP}$ newborn mice. Arrows indicate nucleated erythrocytes. Original magnification: $\times 40$ (**a**, **c**, **d**), $\times 4$ (**b**), $\times 100$ (**e**).

neurogenesis and erythropoiesis, *Mox2*^{+/-cre} *Rb*^{-loxP} mice did exhibit marked defects in other tissues such as the lens and skeletal muscle (A.d.B. and G.L., unpublished observations).

Previous studies have established an essential role for Rb during embryogenesis¹⁶. Here, we find that the cellular architecture of the labyrinth in *Rb*-deficient placenta is markedly altered, resulting in a decreased surface area required for oxygen/nutrient exchange between the mother and fetus. Tetraploid or genetic rescue analysis indicates that a defect in proliferation and differentiation in *Rb*^{-/-} placental trophoblast cells led to the early lethality of *Rb*^{-/-} animals because *Rb*-null embryos, when supplied with a functionally normal placenta, could be carried to full term. Although these experiments define a key role for Rb in the placenta that includes trophoblast stem cells, it is possible that other extra-embryonic cell types may also contribute to the *Rb* knockout phenotype. In this regard, both tetraploid aggregation and conditional knockout approaches restore Rb function in trophoblast cells of the placenta and in the extra-embryonic endoderm of the yolk sac. However, because yolk sac defects manifest as early post-implantation phenotypes, whereas labyrinth defects in the placenta often lead to embryonic lethality after mid-gestation^{17,18}, our current data suggest that the lethality of *Rb*-null embryos at E14.5 is probably due to placental defects. It will be interesting to determine whether loss of Rb in extra-embryonic lineages alone is sufficient to induce embryonic lethality and/or neurogenic and erythroid defects in wild-type fetuses.

So far, rescue of the early embryonic lethality of *Rb*-null animals has been achieved by chimaera approaches^{19–21}, the introduction of an *Rb* minigene²², concomitant mutation of its downstream targets E2F1 (ref. 23), E2F3 (ref. 24) and Id2 (ref. 25), or by reconstitution of a normal placenta (this study). Although in all the previous studies a rescue was achieved by genetically manipulating cells in both embryonic and extra-embryonic tissues^{19–25}, in our study manipulations were confined to extra-embryonic tissues. Our studies suggest that restoration of Rb function in extra-embryonic lineages is sufficient to rescue many of the embryonic defects of *Rb* knockout fetuses, providing a mechanism for the non-cell autonomous roles of Rb during development. Data from the present investigation will stimulate a re-evaluation of numerous previous findings that were based on the use of *Rb* knockout animals and cells. □

Methods

Mouse strains and genotyping

The *Rb*^{+/-} mice³, the *Mox2*^{+/-cre} transgenic mice¹⁴ and the *Rb*^{loxP/loxP} conditional knockout mice¹³ were maintained on a mixed (C57BL/6 × 129/Sv or FVB/N × 129/Sv) background. Genotypes of mice were determined by PCR as previously described^{3,13,14}. Southern blot analysis was carried out on *EcoRV*/*NdeI*-digested genomic DNA using a standard method.

Tetraploid aggregation

Wild-type tetraploid embryos were generated by electrofusion²⁶ of E1.5 embryos derived from crosses between FVB males and ICR females. Fused embryos were cultured overnight in KSOM medium. Diploid embryos were generated from *Rb*^{+/-} intercrosses and were isolated at E2.5 (morula stage). The zona pellucida was removed from all embryos with acidic tyrode solution before aggregation²⁷. Each diploid embryo was aggregated with two tetraploid embryos of the same age (E2.5). Aggregates were cultured overnight and were transferred into the uterus of E2.5 pseudo-pregnant outbred foster females.

Histological analysis and immunohistochemistry

Freshly collected placenta from dated pregnancies were fixed in 4% paraformaldehyde overnight at 4 °C. Neonates were fixed in 10% formalin. Fixed samples were embedded in paraffin and were cut as 5-μm serial sections, which were stained with haematoxylin and eosin for general histopathological analysis.

In situ hybridization and histochemistry were performed on 7-μm histological sections as previously described^{28,29}. For immunohistochemistry, BrdU (100 μg per gram of body weight) was injected intraperitoneally into timed pregnant females 1.5 h before they were killed and the tissue was collected. BrdU-positive cells were detected using a peroxidase-conjugated antibody with the ABC detection system (Santa Cruz). Apoptosis was assayed using the TUNEL/oxidase assay kit (Intergen). Multiple non-consecutive sections at an interval of at least 40 μm were counted for both immunohistochemical stainings and the TUNEL assay. For the BrdU assay, nuclei were counter-stained by haematoxylin. Red blood cells and endothelial cells were excluded from counting. Approximately 400 nuclei

were counted for each placental section. Data were reported as mean percentages (for BrdU) or average numbers (for TUNEL) of positive cells over at least three non-consecutive sections with error bars representing standard deviations of different sections for each sample.

Received 17 October 2002; accepted 7 January 2003; doi:10.1038/nature01417.

1. Nevins, J. R. The Rb/E2F pathway and cancer. *Hum. Mol. Genet.* **10**, 699–703 (2001).
2. Lee, E. Y. *et al.* Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis. *Nature* **359**, 288–294 (1992).
3. Jacks, T. *et al.* Effects of an Rb mutation in the mouse. *Nature* **359**, 295–300 (1992).
4. Clarke, A. R. *et al.* Requirement for a functional Rb-1 gene in murine development. *Nature* **359**, 328–330 (1992).
5. Zhang, P., Wong, C., DePinho, R. A., Harper, J. W. & Elledge, S. J. Cooperation between the Cdk inhibitors p27(KIP1) and p57(KIP2) in the control of tissue growth and development. *Genes Dev.* **12**, 3162–3167 (1998).
6. Faria, T. N., Ogren, L., Talamantes, F., Linzer, D. I. H. & Soares, M. J. Localization of placental lactogen-I in trophoblast giant cells of the mouse placenta. *Biol. Reprod.* **44**, 327–331 (1991).
7. Linzer, D. I., Lee, S. J., Ogren, L., Talamantes, F. & Nathans, D. Identification of proliferin mRNA and protein in mouse placenta. *Proc. Natl Acad. Sci. USA* **82**, 4356–4359 (1985).
8. Lescisin, K. R., Varmuza, S. & Rossant, J. Isolation and characterization of a novel trophoblast-specific cDNA in the mouse. *Genes Dev.* **2**, 1639–1646 (1988).
9. Russ, A. P. *et al.* Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95–99 (2000).
10. Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072–2075 (1998).
11. Metcalfe, L. D., Schmitz, A. A. & Pelka, J. R. Rapid preparation of fatty acid esters from lipids for gas chromatographic analysis. *Anal. Chem.* **38**, 514–515 (1966).
12. James, R. M., Klerkx, A. H., Keighren, M., Flockhart, J. H. & West, J. D. Restricted distribution of tetraploid cells in mouse tetraploid ↔ diploid chimaeras. *Dev. Biol.* **167**, 213–226 (1995).
13. Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev.* **14**, 994–1004 (2000).
14. Tallquist, M. D. & Soriano, P. Epiblast-restricted Cre expression in MORE mice: a tool to distinguish embryonic vs. extra-embryonic gene function. *Genesis J. Genet. Dev.* **26**, 113–115 (2000).
15. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
16. Lipinski, M. M. & Jacks, T. The retinoblastoma gene family in differentiation and development. *Oncogene* **18**, 7873–7882 (1999).
17. Copp, A. J. Death before birth: clues from gene knockouts and mutations. *Trends Genet.* **11**, 87–93 (1995).
18. Cross, J. C., Werb, Z. & Fisher, S. J. Implantation and the placenta: key pieces of the development puzzle. *Science* **266**, 1508–1518 (1994).
19. Lipinski, M. M. *et al.* Cell-autonomous and non-cell-autonomous functions of the Rb tumour suppressor in developing central nervous system. *EMBO J.* **20**, 3402–3413 (2001).
20. Maandag, E. C. *et al.* Developmental rescue of an embryonic-lethal mutation in the retinoblastoma gene in chimeric mice. *EMBO J.* **13**, 4260–4268 (1994).
21. Williams, B. O. *et al.* Extensive contribution of Rb-deficient cells to adult chimeric mice with limited histopathological consequences. *EMBO J.* **13**, 4251–4255 (1994).
22. Zacksenhaus, E. *et al.* pRb controls proliferation, differentiation, and death of skeletal muscle cells and other lineages during embryogenesis. *Genes Dev.* **10**, 3051–3064 (1996).
23. Tsai, K. Y. *et al.* Mutation of E2F-1 suppresses apoptosis and inappropriate S phase entry and extends survival of Rb-deficient mouse embryos. *Mol. Cell* **2**, 293–304 (1998).
24. Ziebold, U., Reza, T., Caron, A. & Lees, J. A. E2F3 contributes both to the inappropriate proliferation and to the apoptosis arising in Rb mutant embryos. *Genes Dev.* **15**, 386–391 (2001).
25. Lasorella, A., Nosedà, M., Beyna, M., Yokota, Y. & Iavarone, A. Id2 is a retinoblastoma protein target and mediates signalling by Myc oncoproteins. *Nature* **407**, 592–598 (2000).
26. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
27. Hogan, B., Beddington, R., Costantini, F. & Lacy, E. *Manipulating the Mouse Embryo: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, Plainview, New York, 1994).
28. Kurz, H., Zechner, U., Orth, A. & Fundele, R. Lack of correlation between placenta and offspring size in mouse interspecific crosses. *Anat. Embryol.* **200**, 335–343 (1999).
29. Anson-Cartwright, L. *et al.* The glial cells missing-1 protein is essential for branching morphogenesis in the chorioallantoic placenta. *Nature Genet.* **25**, 311–314 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank A. Berns, P. Soriano and T. Jacks for providing *Rb*^{+/-loxP}, *Mox2*^{+/-cre} and *Rb*^{+/-} mice, respectively. We also thank S. Hoshaw-Woodard and R. J. Jandacek for technical assistance, and C. Timmers for critical scientific discussions. This work was supported by grants from the National Cancer Institute (to G.L. and T.J.R.), the National Institute of Health (to M.W. and G.L.), the National Center for Research Resources (to T.J.R.), and by the Canadian Institutes of Health Research (to J.C.C.). L.W., H.S. and A.T. were supported by NIH awards, A.d.B. was supported by a UOR Human Cancer Genetics Postdoctoral Fellowship, and G.L. is a V-Foundation and PEW Scholar.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.L. (e-mail: Leone-1@medctr.osu.edu).

Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*

Cas Kramer*, Jennifer J. Loros†‡, Jay C. Dunlap‡ & Susan K. Crosthwaite*

* School of Biological Sciences, University of Manchester, Manchester M13 9PT, UK

† Department of Biochemistry, and ‡ Department of Genetics, Dartmouth Medical School, Hanover, New Hampshire 03755, USA

The prevalence of antisense RNA in eukaryotes is not known and only a few naturally occurring antisense transcripts have been assigned a function^{1–4}. However, the recent identification of a large number of putative antisense transcripts⁵ strengthens the view that antisense RNAs might affect a wider variety of processes than previously thought. Here we show that in the model organism *Neurospora crassa* entrainment of the circadian clock, which is critical for the correct temporal expression of genes and their products, is controlled partly by an antisense RNA arising from a clock component locus. In a wild-type strain, levels of antisense *frequency* (*frq*) transcripts cycle in antiphase to sense *frq* transcripts in the dark, and are inducible by light. In mutant strains in which the induction of antisense *frq* RNA by light is abolished, the time of the internal clock is delayed relative to the wild-type strain, and resetting of the clock by light is altered. These data provide an unexpected link between antisense RNA and circadian timing and provide a new example of a eukaryotic cellular process regulated by naturally occurring antisense RNA.

Circadian timing often involves transcription/translation feedback loops composed of core clock molecules^{6–8}. In *Neurospora* the *frequency* (*frq*), *white collar-1* (*wc-1*) and *white collar-2* (*wc-2*) loci encode important components of the circadian clock^{9–12}. WC-1 and

WC-2 proteins act to increase levels of *frq* messenger RNA^{11,13,14}. After translation, FRQ protein feeds back to downregulate its own expression by means of interactions with the white collar complex^{15–17}. Cyclic expression of *frq*, not just its presence, is a prerequisite for rhythmicity⁹. The presence of antisense *frq* transcripts was first inferred from the sequences of two complementary DNA clones¹⁸. Using riboprobes spanning the length of the sense *frq* transcript, we identified two large antisense *frq* transcripts of ~5 and 5.5 kilobases (kb) in wild-type RNA (Fig. 1a), similar in size to the 4.5-kb and 5-kb sense *frq* mRNA transcripts (Fig. 1a, ref. 20). These transcripts are not detected in a *frq* deletion strain (*frq*¹⁰) (Fig. 1a). When this *frq*-null strain is transformed with an 8.7-kb fragment of the *frq* locus targeted to an ectopic location (*frq*¹⁰KAJ120; see Methods), the wild-type phenotype is rescued¹⁸ and full-length sense and antisense *frq* transcripts are detected (Fig. 1a). This indicates that the antisense *frq* promoter lies within the 3' end of the 8.7-kb fragment (Fig. 1b). We mapped the 3' end of sense and 5' end of antisense *frq* transcripts (Fig. 1b) and found that multiple sense and antisense *frq* transcripts were produced. Computer analyses of the *frq* locus revealed 20 small open reading frames (ORFs) (50–199 amino acids) in the antisense direction, but none shared significant homology with known ORFs, nor were there good consensus sequences for translational initiation. Nevertheless, the detection of both sense and antisense *frq* transcripts in another filamentous fungus, *Chromocrea spinulosa* (C.K. and S.K.C., unpublished observations), indicates conservation consistent with a functional role for this apparently non-coding antisense RNA.

Because of the importance of *frq* gene products in the circadian clock we followed *frq* transcript expression over 2 days when cultures were grown in constant darkness at 25 °C. Under these conditions WC-1 and WC-2 protein activity results in the cyclic expression of *frq* mRNA. As shown previously, levels of sense *frq* RNA cycle in constant darkness (Fig. 1c and d; ref. 9). Antisense *frq* RNA is also rhythmically produced, but 180° out of phase with the sense *frq* RNA (Fig. 1c and d). Peak levels of antisense *frq* RNA in constant darkness are about equal to trough levels of sense *frq* RNA (Fig. 1c). Levels of sense *frq* RNA change up to 9-fold over the course

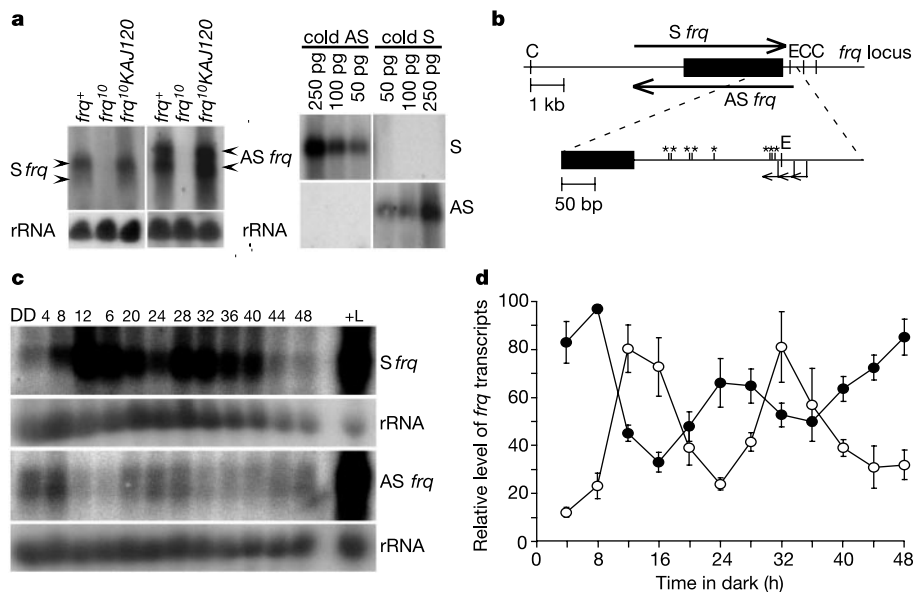


Figure 1 Expression of sense (S) and antisense (AS) RNA transcripts from the *N. crassa* *frq* locus. **a**, Northern analysis of *frq* transcripts (arrowheads) in light-pulsed cultures of wild-type (*frq*⁺), deletion mutant (*frq*¹⁰), 'rescued wild-type' (*frq*¹⁰KAJ120) and 'cold controls' (see Methods). **b**, Schematic representation of the *frq* locus, contained within an 8.7-kb *Cla*I fragment²⁰. Filled boxes, FRQ ORF; asterisks, polyadenylation sites of sense *frq* mRNA (mapped to 54, 57, 85, 90, 117, 201, 202, 205 and 208 bp downstream

of the stop codon of the FRQ ORF); small arrows, transcription start points of antisense *frq* transcripts (mapped to 217, 241 and 278 bp downstream of the stop codon of the FRQ ORF); C, *Cla*I; E, *Eco*RV. **c**, Rhythmic expression of *frq* transcripts in constant darkness (DD) and after a light pulse (+L). **d**, Densitometric analysis of *frq* transcript levels (mean ± s.e.m., *n* = 5). Highest value set to 100%. Open circles, sense *frq* RNA; filled circles, antisense *frq* RNA.

of a circadian day; in comparison, the amplitude of cycling antisense *frq* RNA is only ~2–2.5-fold (Fig. 1d). Like sense *frq* RNA, both antisense *frq* transcripts accumulate markedly in response to light (Fig. 1c). On transfer to the dark, *frq* RNA decreases rapidly to a low level¹⁹. Because *frq* expression is autoregulatory⁹, one explanation for this rapid decrease is its transcriptional repression on transfer to constant darkness owing to high levels of light-induced FRQ protein¹⁹. However, it is also possible that the half-life of the sense *frq* transcript is decreased as a result of the action of antisense *frq* RNA.

To investigate the role of antisense *frq* transcripts we constructed a mutant strain, *frq*¹⁰*frqccg-2*, in which the 3' end of the *frq* locus was replaced with 3' untranslated region (UTR) of a transcriptionally controlled *Neurospora* gene, *eas* (*cgc-2*)^{20,21} (Fig. 2a). We postulated that in the *frq*¹⁰*frqccg-2* strain the expression of antisense *frq* RNA would be abolished owing to the removal of its promoter sequences, whereas the expression of sense *frq* RNA would be

unaltered. The expression pattern of sense and antisense *frq* RNA in strain *frq*¹⁰*frqccg-2* compared with *frq*⁺ or *frq*¹⁰*KAJ120* after a transfer from light to dark is indeed markedly different. In the *frq*¹⁰*frqccg-2* strain, expression of antisense *frq* transcripts in the dark is considerably decreased and the induction of antisense *frq* RNA in the light is eliminated (Fig. 2b). Overall levels of sense *frq* are lower in *frq*¹⁰*frqccg-2*, and peak expression of the sense *frq* transcripts is delayed by ~4 h (Fig. 2b and c).

To study the physiological consequence of altered *frq* expression we assayed the clock-controlled switch between vegetative hyphal growth and asexual spore production (conidiation), which is classically used as a means to monitor the state of the circadian clock²⁰. Between 18 and 30 °C conidiation in the wild-type strain is rhythmic²⁰. At the lower end of *Neurospora*'s physiological growth range (18 and 20 °C) only the first band of conidiation is formed in *frq*¹⁰*frqccg-2*, after which conidiation is arrhythmic. At higher temperatures (22–30 °C) rhythmic conidiation is observed, and at 25 and 28 °C there was a small but significant lengthening of the period compared with *frq*¹⁰*KAJ120* (at 25 °C, 0.8 ± 0.1 h; at 28 °C, 0.7 ± 0.1 h; $P < 0.01$). Moreover, in the *frq*¹⁰*frqccg-2* strain, which lacks light-induced antisense *frq* RNA, the transfer from light to dark sets the clock to a time 1–3 h earlier than in the *frq*⁺ strain (Table 1). This is seen as a significant delay in the onset of conidiation (Fig. 3a) and is correlated with the delayed peak in sense *frq* expression in this mutant strain (Fig. 2b and c). One reason for this might be that, in the absence of antisense *frq* transcripts, levels of sense *frq* RNA remain higher after a transfer from light to dark. In this situation a longer time might have to elapse before levels of sense *frq* RNA and thus FRQ protein fall sufficiently to initiate the cyclic expression of *frq* mRNA. However, we found that the delayed increase of sense *frq* RNA was not correlated with elevated levels of sense *frq* RNA in *frq*¹⁰*frqccg-2* after transfer from light to dark (Fig. 3b). Nonetheless, the molecular and physiological phenotypes indicate that the antisense *frq* transcripts are important in synchronizing internal to external time. This was examined further.

In all organisms, exposure to light in the early evening delays the time of the circadian clock, whereas exposure to light late in the night advances the time of the clock²². Both sense and antisense *frq* transcripts are induced to their highest levels after a light pulse, and induction of sense *frq* by light is required for light-resetting of the *Neurospora* clock¹⁹. We therefore assessed the effect of light pulses on clock resetting at different times of the circadian day. If antisense *frq* were involved in clock function, perhaps by regulating expression of sense *frq* transcripts, this assay would reveal altered responses of the antisense-defective strain (*frq*¹⁰*frqccg-2*) to light. In Fig. 4a the response of *frq*⁺, *frq*¹⁰*KAJ120* and *frq*¹⁰*frqccg-2* to 2-min light pulses given to dark-grown cultures at 2-h intervals over one circadian day is shown. Under these conditions only small advances and delays in the onset of conidiation are seen in the *frq*⁺ and *frq*¹⁰*KAJ120* strains. In contrast, in the *frq*¹⁰*frqccg-2* strain the

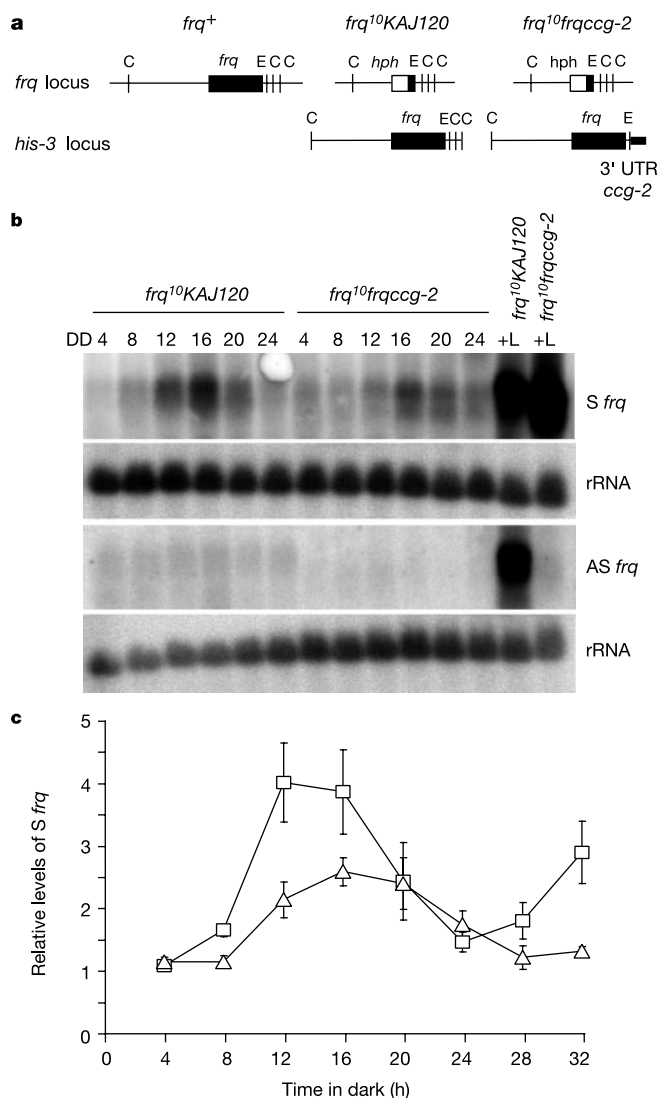


Figure 2 Altered expression of sense (S) and antisense (AS) *frq* transcripts in mutant strain *frq*¹⁰*frqccg-2*. **a**, Schematic representation of the *frq* and *his-3* loci of *frq*⁺, *frq*¹⁰*KAJ120* and *frq*¹⁰*frqccg-2*: thick filled boxes, FRQ ORF; open boxes, hygromycin phosphotransferase (*hph*) sequences; thin filled box, 3' UTR sequences of *cgc-2*, *Clal*; E, *EcoRV*. **b**, Levels of *frq* transcripts at 25 °C in constant darkness (DD) and 30 min after a 2-min saturating light pulse at DD24 (+L). **c**, Densitometric analysis of sense *frq* transcript levels (mean ± s.e.m., $n = 4$ or 5). Lowest value set to 1.0. Squares, *frq*¹⁰*KAJ120*; triangles, *frq*¹⁰*frqccg-2*.

Table 1 Circadian time of *frq*¹⁰*frqccg-2* after transfer from light to dark

Temperature (°C)	Circadian time (h)		
	<i>frq</i> ⁺	<i>frq</i> ¹⁰ <i>KAJ120</i>	<i>frq</i> ¹⁰ <i>frqccg-2</i>
18	12.5 ± 0.4	14.1 ± 1.1	14.0 ± 0.4
20	14.8 ± 1.2	14.4 ± 1.6	15.9 ± 0.2
22	11.9 ± 0.4	14.3 ± 0.5	12.8 ± 0.3
24	13.4 ± 0.5	13.7 ± 0.2	11.3 ± 0.6
25	12.3 ± 0.3	13.6 ± 0.2	11.3 ± 0.2
28	12.1 ± 0.3	13.6 ± 0.3	10.5 ± 0.3
30	12.0 ± 0.2	11.4 ± 0.3	8.5 ± 0.3

The light-to-dark transfer sets the clock to a different circadian time in *frq*⁺, *frq*¹⁰*KAJ120* and *frq*¹⁰*frqccg-2* (see also Methods). For phase calculations the first band of conidiation in 4–12 individual race tubes was used. The centre of the conidial band was used as a phase reference point and was set to circadian time 0. Phase differences between *frq*⁺ and *frq*¹⁰*frqccg-2* and between *frq*¹⁰*KAJ120* and *frq*¹⁰*frqccg-2* at 24, 25, 28 and 30 °C are statistically different ($P < 0.01$). Results are means ± s.e.m.

clock resets by up to 12 h. Thus, our data indicate that the absence of light-induced antisense *frq* RNA has a marked effect on the clock's response to light.

One caveat of this conclusion is that the phenotype of *frq*¹⁰*frqccg-2* might be due to an alteration of 3' UTR sequences of the sense transcript. We therefore mapped the 3' ends of sense *frq* transcripts in *frq*¹⁰*frqccg-2* and found that most are chimaeric because the polyadenylation signal in the *cgc-2* sequence is used predominantly, resulting in a 3' UTR of 335 bp (222 bp of *frq* sequences and 113 bp of *cgc-2* sequences). We therefore constructed a second mutant strain, *frq*¹⁰*KAJ128*, identical to *frq*¹⁰*KAJ120* except that it lacked the final 472 bp of the 8.7-kb *frq* sequence (Fig. 4b). In *frq*¹⁰*KAJ128* several different sense 3' ends were mapped, corresponding in sequence and abundance to those in the wild-type *frq*⁺ strain (50, 80, 90 and 208 bp downstream of the FRQ ORF; see Fig. 1b). In contrast to *frq*¹⁰*frqccg-2*, levels of antisense *frq* transcripts present in the dark in *frq*¹⁰*KAJ128* are comparable to those in the wild type. However, as with *frq*¹⁰*frqccg-2*, the expression of antisense *frq* RNA is not light-induced in *frq*¹⁰*KAJ128* (Fig. 4b, and data not shown). A sequence similar to that of a light-responsive element in the sense *frq* promoter¹⁴ is present in the antisense promoter sequences deleted from *frq*¹⁰*KAJ128*. Nevertheless, in response to light, sense *frq* RNA levels rise with similar kinetics in *frq*¹⁰*KAJ120*, *frq*¹⁰*frqccg-2* and *frq*¹⁰*KAJ128* (Fig. 4b) and FRQ protein levels, and phosphorylation status seemed normal in protein extracted from light-grown and dark-grown *Neurospora* cultures (data not shown). When the light-pulse experiment was repeated with *frq*¹⁰*KAJ128* the same marked response to light pulses given throughout the day was seen, as observed for *frq*¹⁰*frqccg-2* (Fig. 4c). These data indicate that it is indeed the lack of light-induced antisense *frq* RNA, rather than an altered sense *frq* RNA, that is responsible for the spectacular resetting of the *Neurospora* clock in response to light in *frq*¹⁰*frqccg-2* and *frq*¹⁰*KAJ128*. The response of the clock in these strains to light pulses was similar to that of a *Neurospora* vivid mutant²³; however, VIVID gates the response of a number of

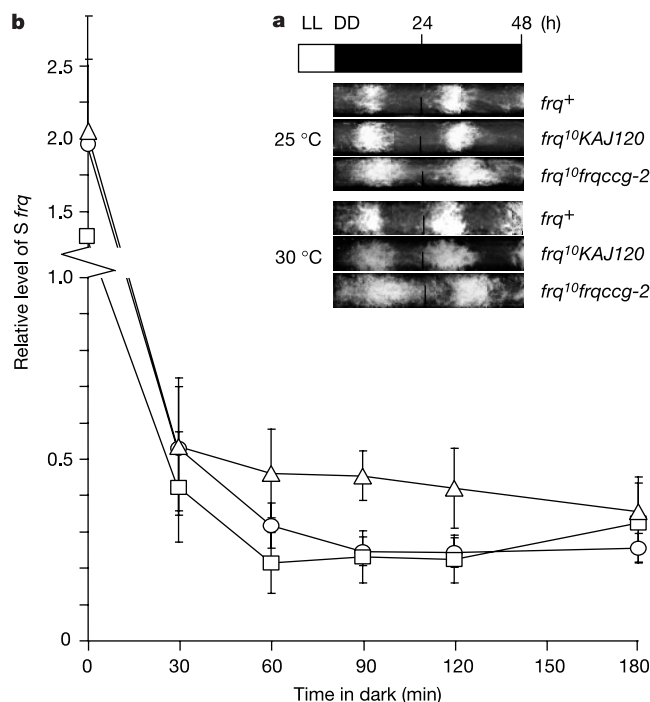


Figure 3 Circadian and molecular phenotype of *frq*¹⁰*frqccg-2* after a light (LL) to dark (DD) transfer. **a**, Phenotype of *frq*⁺, *frq*¹⁰*KAJ120* and *frq*¹⁰*frqccg-2* grown on race tubes; 48 h growth is shown. **b**, Densitometric analysis of sense *frq* transcript levels (mean $\pm$ s.e.m., $n = 3$). The intensity of the hybridization signal of 100 pg of 'cold control' was set at 1.0. Circles, *frq*⁺; squares, *frq*¹⁰*KAJ120*; triangles, *frq*¹⁰*frqccg-2*.

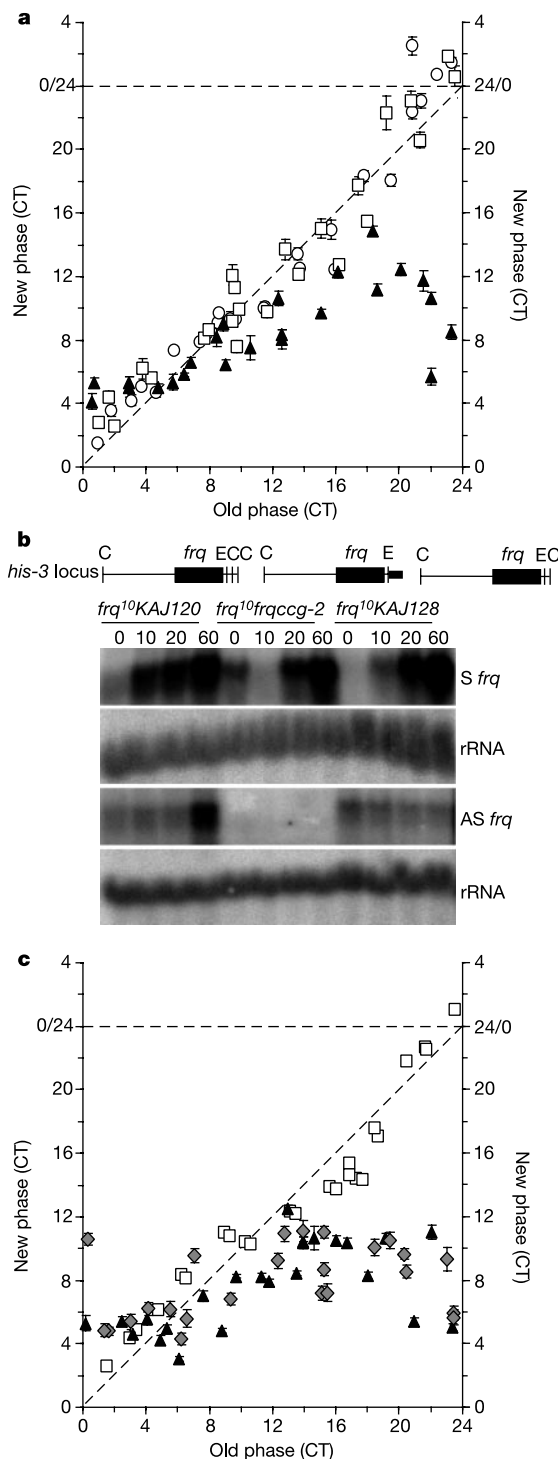


Figure 4 The response of *frq*¹⁰*frqccg-2* and *frq*¹⁰*KAJ128* to light pulses. **a**, **c**, The resetting response of the clock to a 2-min saturating light pulse given at intervals throughout the circadian day is plotted as the circadian time (CT) before (old phase) against the circadian time after the light pulse (new phase). The dashed diagonal line represents the 'no response line'; data points above this line represent clock delays, whereas data points below this line represent clock advances. Open circles, phase response data of *frq*⁺, open squares, *frq*¹⁰*KAJ120*, filled triangles, *frq*¹⁰*frqccg-2*; filled diamonds, *frq*¹⁰*KAJ128*. Results were calculated from four to six race tubes (mean $\pm$ s.e.m.) in four separate experiments (two experiments are shown in **a** and two in **c**). The phase of the second and third bands of conidiation after the light pulse was used to calculate new phase. **b**, Schematic representation of the *his-3* loci of *frq*¹⁰*KAJ120*, *frq*¹⁰*frqccg-2* and *frq*¹⁰*KAJ128* (see also Fig. 2a) and levels of sense (S) and antisense (AS) *frq* transcripts 15 min after exposure to 0, 10, 20 and 60 s of dim light (0.4 μ mol m⁻² s⁻¹ = 33 lux) at CT21.

different photo-controlled processes to light, and in its absence mutants are bright orange owing to the overproduction of carotenoids. In contrast, neither *frq¹⁰frqccg-2* nor *frq¹⁰KAJ128* show signs of increased carotenoid synthesis, perhaps indicating that the effect of antisense *frq* RNA is clock-specific.

The almost totally overlapping nature of sense and antisense *frq* transcripts is intriguing and might reflect the importance of complementarity or convergent transcription for the regulation or function of antisense *frq* RNA. The strong resetting of the clock by light in the absence of light-induced antisense *frq* transcripts indicates that the activity of at least one clock component has been altered in *frq¹⁰frqccg-2* and *frq¹⁰KAJ128*. Moreover, in dark-grown cultures of *frq¹⁰frqccg-2* and *frq¹⁰KAJ128* the delayed increase of sense *frq* transcripts and significantly lower peak levels implicate an effect of antisense *frq* RNA on the transcription or stability of sense *frq* RNA. As there is no evidence that antisense *frq* RNA encodes a protein, it might affect the expression of sense *frq* or indeed other clock components by means of the formation of RNA–RNA or RNA–protein complexes. Post-transcriptional regulation by small interfering antisense RNAs, deriving from double-stranded RNA, has been reported in a wide range of organisms including *Neurospora*²⁴. Although the mechanism is not completely understood, the ribonuclease Dicer is required²⁵. Dicer orthologues are present in *Neurospora*, and if *frq* RNA is a target the resulting small RNAs could regulate sense *frq*, by altering either its stability or its availability for translation. However, because levels of sense *frq* RNA in constant light are similar regardless of antisense *frq* levels, an effect on *frq* RNA stability in constant light is not supported by our data. Nevertheless, it is possible that immediately after a light pulse given to dark-grown cultures, sense *frq* RNA stability or translation is affected in the antisense-defective strains. Alternatively, antisense *frq* RNA might affect protein activity. Although FRQ protein levels and phosphorylation status seem normal in light-grown and dark-grown cultures of *frq¹⁰frqccg-2* and *frq¹⁰KAJ128*, post-translational modifications of FRQ, or other clock components, that could not be detected in our assays might have occurred. Induction of antisense *frq* RNA by light is, like that of sense *frq*, abolished in a *wc-1* mutant strain (data not shown), indicating that antisense *frq* expression might be activated and repressed in a similar manner to that of sense *frq* in response to at least one environmental signal¹¹. If both sense and antisense transcripts are regulated by similar environmental factors, one role of antisense *frq* RNA might be to limit the response of the clock to extremes in the environment, thus conferring the ability to keep accurate time. In the same way, other antisense RNAs might be involved in the maintenance of homeostasis. In the silkworm, *Antheraea pernyi*, an endogenous antisense transcript with homology to the *Drosophila* clock component *period* (*per*) has also been reported²⁶; however, sense and antisense *per* transcripts are encoded on different chromosomes; antisense *per* RNA originates from the W chromosome, so any function must be female-specific²⁷. Although we cannot yet discuss the exact mechanism by which the antisense *frq* transcripts exert their effects on the clock, these data are clear in showing a role for antisense regulation in circadian timing. Antisense *frq* RNA is expressed at a physiologically significant level and is regulated by appropriate environmental stimuli; in addition, a loss of antisense *frq* transcripts reveals a role for its action in setting the phase of the rhythm. □

Methods

Fungal strains

N. crassa *frq⁺* 'lab wild-type' [30–7 *bd*; *frq⁺* A], deletion mutant *frq¹⁰* [93–4 *bd*; *his-3*; *frq¹⁰* A] and *frq¹⁰KAJ120* [93–4 *bd*; *his-3* (*his-3⁺*::*pKAJ120*); *frq¹⁰* A] have been described previously^{9,18}. *N. crassa* strains *frq¹⁰KAJ120* [93–4 *bd*; *his-3* (*his-3⁺*::*pKAJ120*); *frq¹⁰* A], *frq¹⁰frqccg-2* [93–4 *bd*; *his-3* (*his-3⁺*::*pfrqccg-2*); *frq¹⁰* A] and *frq¹⁰KAJ128* [93–4 *bd*; *his-3* (*his-3⁺*::*pKAJ128*); *frq¹⁰* A] were constructed as follows. Deleting an 897-bp *EcoRV*–*Clal* fragment from plasmid *pKAJ120* (ref. 28) and replacing this fragment with a 520-bp *EcoRV*–*XbaI*^{blunt} fragment containing 3' UTR sequences from *N. crassa eas* (*cgc-2*) created the plasmid *pfrqccg-2*. Plasmid *pKAJ128* was created by deleting a 472 bp *Clal*^{methyated}–

Clal fragment from *pKAJ120*. Strains *frq¹⁰KAJ120*, *frq¹⁰frqccg-2* and *frq¹⁰KAJ128* were obtained by site-directed integration of plasmids *pKAJ120*, *pfrqccg-2* and *pKAJ128*, respectively, into the *his-3* locus of the *frq¹⁰* deletion mutant as previously described¹⁸. The integrity of independently isolated transformants of these strains was confirmed by Southern analysis and polymerase chain reaction (PCR).

Race tube assays

Standard 30-cm 'race tubes' (containing 0.1% glucose, 0.17% arginine, 1 × Vogel's salts, biotin (0.5 mg l⁻¹), 1.5% agar) were used for analysis of asexual sporulation (conidiation). Strains were inoculated and grown in constant light at 25 °C for 24 h before transfer to constant darkness at different experimental temperatures. For densitometric analyses the CHRONO program²⁹ was used. Circadian time (CT) is derived by dividing the period of the rhythm into 24 equal parts. By convention CT0 is subjective dawn and CT12 is subjective dusk²⁰.

3' RACE and 5' RACE

The 3' RACE System and the 5' RACE System (Invitrogen) were used essentially in accordance with the manufacturer's recommendations. For PCR amplification *Taq* DNA polymerase (Roche) and HotStarTaq DNA polymerase (Qiagen) were used with the gene-specific primers 5' bgl2-F (5'–GCTGGGACGATGGCGATGATCT–3'), 3' bgl-F (5'–GCGCCGAGCAATCTCTCATTTC–3') and 3' bgl-F (5'–GGAAGGACGATTTGGCG TTCGATA–3'). Total RNA was isolated with RNeasy kit (Qiagen) or Trizol reagent (Invitrogen).

Northern analysis

Tissue was harvested as described previously¹⁹ and total RNA was extracted with Trizol in accordance with the manufacturer's protocol. Northern blot hybridization was performed as described previously⁹ with the use of [³²P]UTP-labelled riboprobes spanning the full length of the sense *frq* transcripts. Riboprobes transcribed *in vitro* from a PCR amplification product of the 2.2-kb *EcoRI* fragment within the *N. crassa* FRQ ORF with T3 and T7 RNA polymerases (Ambion or MBI Fermentas) were used in all northern analyses. On each northern blot 50–500 pg of sense and antisense 'cold control' transcripts transcribed *in vitro* (from the same 2.2-kb PCR template) were present to confirm the specificity of the northern hybridization bands observed (see also Fig. 1a). Signals from PhosphorImager screens were quantified with the software package QuantityOne (BioRad) and normalized to 18S ribosomal RNA.

Received 22 October 2002; accepted 13 January 2003; doi:10.1038/nature01427.

- Kumar, M. & Carmichael, G. G. Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. *Microbiol. Mol. Biol. Rev.* **62**, 1415–1434 (1998).
- Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.* **2**, 919–929 (2001).
- Avner, P. & Heard, E. X-chromosome inactivation: counting, choice and initiation. *Nature Rev. Genet.* **2**, 59–67 (2001).
- Sleutels, F., Zwart, R. & Barlow, D. P. The noncoding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
- Ambros, V. MicroRNAs: tiny regulators with great potential. *Cell* **107**, 823–826 (2001).
- Dunlap, J. C. Molecular bases for circadian clocks. *Cell* **96**, 271–290 (1999).
- Cermakian, N. & Sassone-Corsi, P. Multilevel regulation of the circadian clock. *Nature Rev. Mol. Cell Biol.* **1**, 59–67 (2000).
- Hastings, M. H. Circadian clockwork: two loops are better than one. *Nature Rev. Neurosci.* **1**, 143–145 (2000).
- Aronson, B. D., Johnson, K. A., Loros, J. J. & Dunlap, J. C. Negative feedback defining a circadian clock: autoregulation of the clock gene *frequency*. *Science* **263**, 1578–1584 (1994).
- Merrow, M. W. & Dunlap, J. C. Intergenic complementation of a circadian rhythmicity defect: phylogenetic conservation of structure and function of the clock gene *frequency*. *EMBO J.* **13**, 2257–2266 (1994).
- Crosthwaite, S. K., Dunlap, J. C. & Loros, J. J. *Neurospora wc-1* and *wc-2*: transcription, photoreponses, and the origins of circadian rhythmicity. *Science* **276**, 763–769 (1997).
- Cheng, P., Yang, Y. & Liu, Y. Interlocked feedback loops contribute to the robustness of the *Neurospora* circadian clock. *Proc. Natl Acad. Sci. USA* **98**, 7408–7413 (2001).
- Collett, M. A., Dunlap, J. C. & Loros, J. J. Circadian clock-specific roles for the light response protein WHITE COLLAR-2. *Mol. Cell Biol.* **21**, 2619–2628 (2001).
- Fröhlich, A. C., Liu, Y., Loros, J. J. & Dunlap, J. C. White Collar-1, a circadian blue light photoreceptor, binding to the *frequency* promoter. *Science* **297**, 815–819 (2002).
- Denault, D. L., Loros, J. J. & Dunlap, J. C. WC-2 mediates WC-1–FRQ interaction within the PAS protein-linked circadian feedback loop of *Neurospora*. *EMBO J.* **20**, 109–117 (2001).
- Merrow, M. et al. Circadian regulation of the light input pathway in *Neurospora crassa*. *EMBO J.* **20**, 307–315 (2001).
- Cheng, P., Yang, Y., Heintzen, C. & Liu, Y. Coiled-coil domain-mediated FRQ–FRQ interaction is essential for its circadian clock function in *Neurospora*. *EMBO J.* **20**, 101–108 (2001).
- Aronson, B. D., Johnson, K. A. & Dunlap, J. C. Circadian clock locus *frequency*: protein encoded by a single open reading frame defines period length and temperature compensation. *Proc. Natl Acad. Sci. USA* **91**, 7683–7687 (1994).
- Crosthwaite, S. K., Loros, J. J. & Dunlap, J. C. Light-induced resetting of a circadian clock is mediated by a rapid increase in *frequency* transcript. *Cell* **81**, 1003–1012 (1995).
- Loros, J. J. & Dunlap, J. C. Genetic and molecular analysis of circadian rhythms in *Neurospora*. *Annu. Rev. Physiol.* **63**, 757–794 (2001).
- Bell-Pedersen, D., Dunlap, J. C. & Loros, J. J. The *Neurospora* circadian clock-controlled gene, *cgc-2*, is allelic to *eas* and encodes a fungal hydrophobin required for formation of the conidial rodlet layer. *Genes Dev.* **6**, 2382–2394 (1992).
- Johnson, C. H. Forty years of PRCs: what have we learned? *Chronobiol. Int.* **16**, 711–743 (1999).
- Heintzen, C., Loros, J. J. & Dunlap, J. C. The PAS protein VIVID defines a clock-associated feedback loop that represses light input, modulates gating, and regulates clock resetting. *Cell* **104**, 453–464 (2001).

24. Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. Involvement of small RNAs and role of the *qde* genes in the gene silencing pathway in *Neurospora*. *Genes Dev.* **16**, 790–795 (2002).
25. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
26. Sauman, I. & Reppert, S. M. Circadian clock neurons in the silkworm *Antheraea pernyi*: novel mechanisms of *period* protein regulation. *Neuron* **17**, 889–900 (1996).
27. Gotter, A. L., Levine, J. D. & Reppert, S. M. Sex-linked period genes in the silkworm, *Antheraea pernyi*: implications for circadian clock regulation and the evolution of sex chromosomes. *Neuron* **24**, 953–965 (1999).
28. Johnson, K. A. *Molecular Characterization of the Circadian Clock Locus Frequency of Neurospora crassa*. Thesis, Dartmouth College (1993).
29. Roenneberg, T. & Taylor, W. Automated recordings of bioluminescence with special reference to the analysis of circadian rhythms. *Methods Enzymol.* **305**, 104–119 (2000).
30. McClung, C. R., Fox, B. A. & Dunlap, J. C. The *Neurospora* clock gene *frequency* shares a sequence element with the *Drosophila* clock gene *period*. *Nature* **339**, 558–562 (1989).

Acknowledgements We thank K. Gull, A. Loudon and S. Oliver for critical reading of the manuscript; C. Heintzen for discussion and critical reading of the manuscript; and J. Miller for technical assistance. This work was supported by grants from the BBSRC, NIGMS, NIMH and NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K.C. (e-mail: susan.k.crosthwaite@man.ac.uk).

MDC1 is required for the intra-S-phase DNA damage checkpoint

Michal Goldberg^{*,†}, Manuel Stucki^{*,†}, Jacob Falck[‡], Damien D'Amours^{*}, Dinah Rahman[§], Darryl Pappin[§], Jiri Bartek[‡] & Stephen P. Jackson^{*}

^{*} The Wellcome Trust/Cancer Research UK Institute of Cancer and Developmental Biology and Department of Zoology, University of Cambridge, Cambridge CB2 1QR, UK

[‡] Institute of Cancer Biology, Danish Cancer Society, Strandboulevard 49, DK-2100, Copenhagen, Denmark

[§] Department of Proteomics, Imperial College School of Medicine, Hammersmith Campus, London W12 0NN, UK

[†] These authors contributed equally to this work

MRE11, RAD50 and NBS1 form a highly conserved protein complex (the MRE11 complex) that is involved in the detection, signalling and repair of DNA damage¹. We identify MDC1 (KIAA0170/NFBD1), a protein that contains a forkhead-associated (FHA) domain and two BRCA1 carboxy-terminal (BRCT) domains, as a binding partner for the MRE11 complex. We show that, in response to ionizing radiation, MDC1 is hyperphosphorylated in an ATM-dependent manner, and rapidly relocates to nuclear foci that also contain the MRE11 complex, phosphorylated histone H2AX and 53BP1. Downregulation of MDC1 expression by small interfering RNA yields a radio-resistant DNA synthesis (RDS) phenotype and prevents ionizing radiation-induced focus formation by the MRE11 complex. However, downregulation of MDC1 does not abolish the ionizing radiation-induced phosphorylation of NBS1, CHK2 and SMC1, or the degradation of CDC25A. Furthermore, we show that overexpression of the MDC1 FHA domain interferes with focus formation by MDC1 itself and by the MRE11 complex, and induces an RDS phenotype. These findings reveal that MDC1-mediated focus formation by the MRE11 complex at sites of DNA damage is crucial for the efficient activation of the intra-S-phase checkpoint.

After exposure of cells to ionizing radiation (IR) or to radio-mimetic drugs, the function of the MRE11 complex is regulated by the phosphorylation of NBS1 by ATM^{2–5}. In humans, mutations in

ATM, MRE11 and NBS1 cause ataxia–telangiectasia (AT), AT-like disorder (AT-LD) and Nijmegen breakage syndrome (NBS), respectively^{6–10}. These disorders are associated with increased predisposition to cancer, and they share similar cellular phenotypes, including chromosomal instability, radiosensitivity and RDS^{6,8,11,12}. The RDS phenotype of NBS and AT-LD cells indicates that the MRE11 complex is crucial for the intra-S-phase checkpoint, where it seems to act parallel to the ATM–CHK2–CDC25A pathway¹³.

To gain further insights into the functions of the MRE11 complex, we raised and affinity-purified antibodies against MRE11 and RAD50, then used these to immuno-purify the MRE11 complex from HeLa cell nuclear extracts. Apart from MRE11, RAD50 and NBS1, we consistently retrieved several other polypeptides, the most prominent being a protein with a relative molecular mass of ~250,000 (*M_r* 250K) (p250; Fig. 1a). Mass spectrometric analyses identified p250 as the product of the human complementary DNA clone KIAA0170 (ref. 14). This protein, henceforth named MDC1 (for mediator of DNA damage checkpoint protein 1), comprises 2,089 amino-acid residues and has a predicted *M_r* of 226.5K. MDC1 contains an FHA domain at its amino terminus and two BRCT domains at its carboxy terminus (Fig. 1b). These modular domains are often found in proteins involved in DNA damage responses and/or cell-cycle control^{15,16}. In addition, the central region of

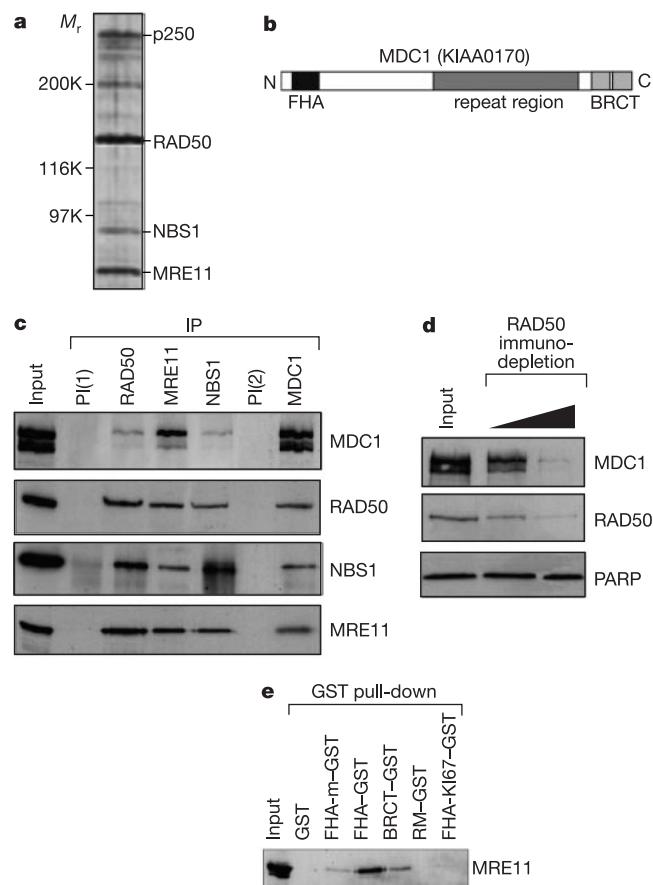


Figure 1 Association of MDC1 with the MRE11 complex. **a**, Identification of MDC1 (p250) in MRE11 immunoprecipitates. **b**, Schematic representation of MDC1. **c**, Co-immunoprecipitation of the MRE11 complex with MDC1. Immunoprecipitations (IP) were carried out and analysed with the indicated antibodies. PI(1) and PI(2) are rabbit and sheep pre-immune sera, respectively. **d**, Immunodepletion of MDC1 by anti-RAD50 antibodies. Supernatants from immunoprecipitations using anti-RAD50 antibodies (25 and 50 μ g) were probed with the indicated antibodies. **e**, The MDC1 FHA domain retrieves MRE11 from nuclear extract. Proteins bound to the GST fusions were probed with anti-MRE11 antibodies.

MDC1 contains 19 consecutive repeats of a motif of ~40 amino-acid residues.

To confirm the interaction between MDC1 and the MRE11 complex, we raised antibodies against various regions of MDC1.

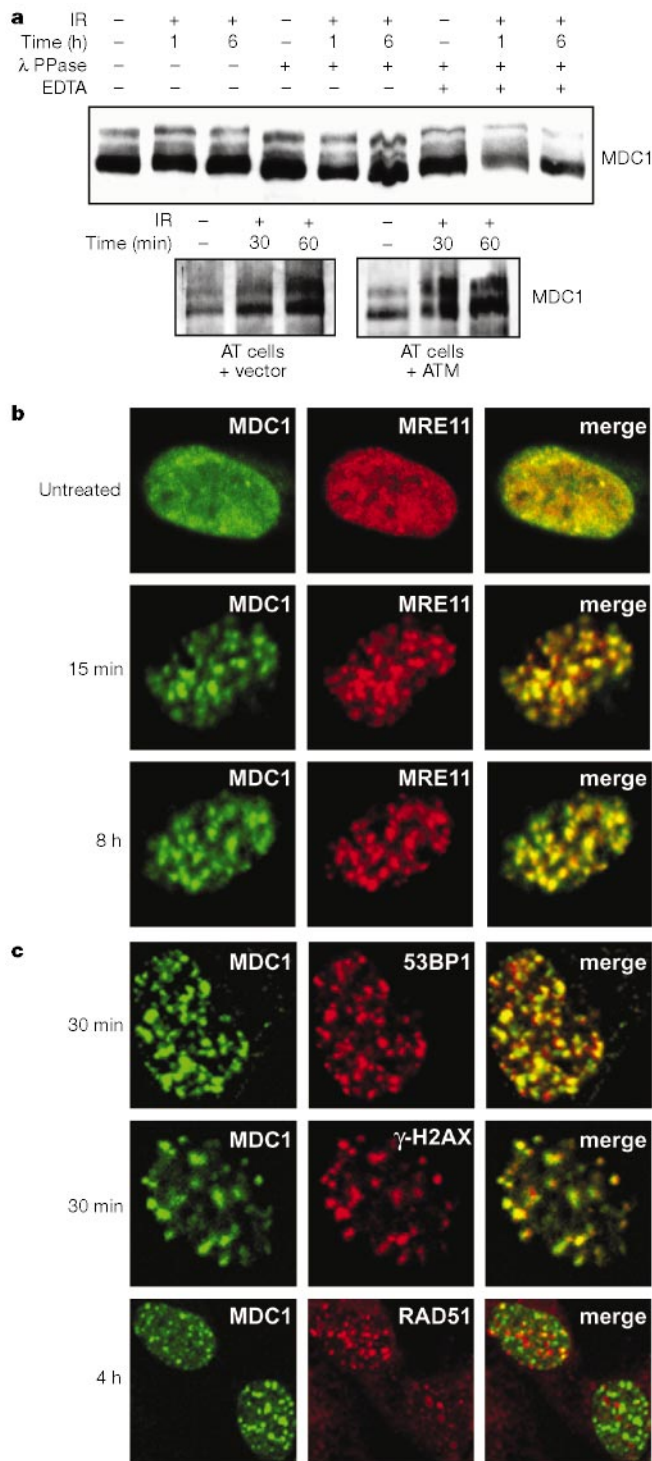


Figure 2 MDC1 responds to ionizing radiation. **a**, MDC1 IR-induced hyperphosphorylation is ATM-dependent. Top, HeLa cell extracts were prepared 1 and 6 h after irradiation (20 Gy). Where indicated, λ phosphatase (λ PPase) and EDTA were added. Bottom, MDC1 was analysed in AT cells and in AT cells complemented with ATM, 30 and 60 min after irradiation (10 Gy). **b**, Colocalization of MDC1 with the MRE11 complex. HeLa cells were analysed at the indicated times after irradiation (10 Gy). **c**, Colocalization of MDC1 with foci formed by 53BP1 and γ -H2AX. Irradiated MRC5 cells were analysed at the indicated times after irradiation.

Each of these detected two to three bands of $M_r > 210$ K in immunoblot analysis of nuclear extracts (Figs 1c, d and 2a), which probably represent alternatively spliced forms of MDC1 because a search of human expressed sequence tag databases indicated the existence of alternatively spliced forms of MDC1 (data not shown). MDC1 was co-immunoprecipitated from HeLa nuclear extracts by antibodies against MRE11, RAD50 or NBS1 (Fig. 1c). Moreover, antibodies against MDC1 efficiently co-immunoprecipitated MRE11, RAD50 and NBS1 (Fig. 1c), indicating that MDC1 interacts with the intact MRE11 complex. Notably, MDC1 could be almost completely depleted from the extract by antibodies against RAD50 (Fig. 1d), showing that most MDC1 is associated with the MRE11 complex under these conditions. The above studies were carried out in the presence of ethidium bromide, indicating that the interaction is not mediated by DNA.

To investigate how MDC1 interacts with the MRE11 complex, we used glutathione S-transferase (GST) fusions of MDC1 domains in *in vitro* binding studies. As shown in Fig. 1e, the MDC1 FHA

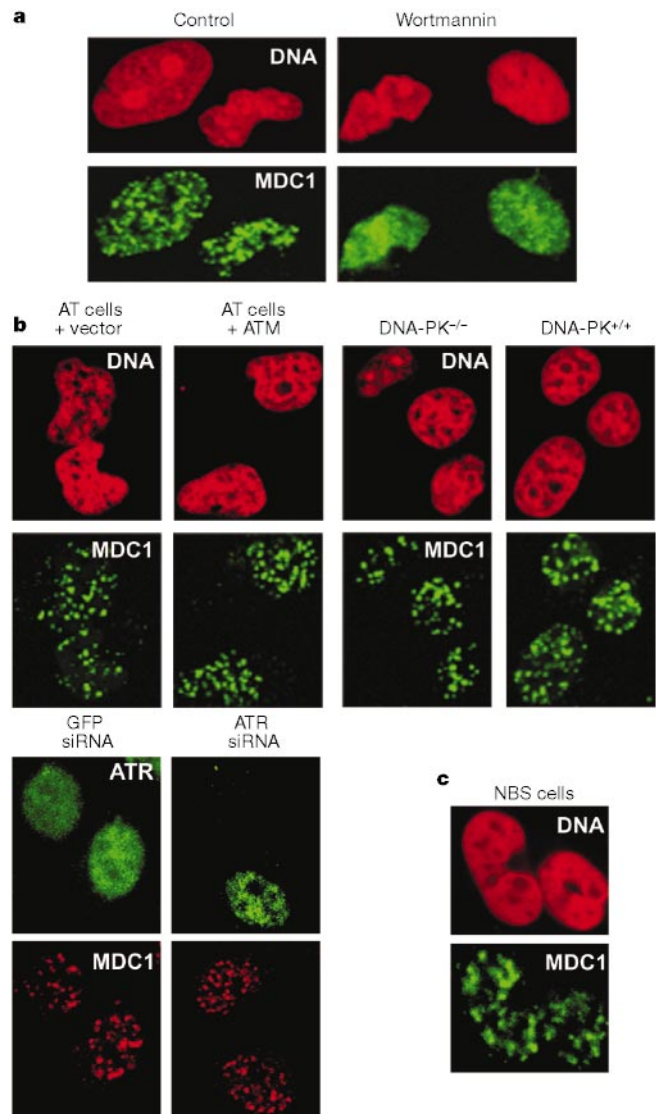


Figure 3 MDC1 focus formation does not require ATM, ATR, DNA-PK or NBS1.

a, Wortmannin inhibits MDC1 focus formation. Cells were pretreated with wortmannin (30 min) and analysed 30 min after irradiation (10 Gy). **b**, MDC1 forms foci in cells that lack ATM, DNA-PK or ATR. DNA-PK^{+/+}, DNA-PK^{-/-}, AT complemented with ATM, and AT cells were analysed 20 min after irradiation (5 Gy). HeLa cells were irradiated (10 Gy) 96 h after ATR siRNA transfection, and analysed 1 h after irradiation. **c**, MDC1 forms foci in NBS fibroblasts. NBS-1LBI cells were analysed 30 min after irradiation (10 Gy).

domain retrieved the MRE11 complex from nuclear extracts. By contrast, a point-mutated derivative of this domain (FHA-m-GST) predicted to be impaired in phospho-dependent protein binding¹⁵, and the MDC1 BRCT domains bound only poorly to the MRE11 complex. The MRE11 complex was not retrieved at all from extracts by a fragment derived from the internal repeat region of MDC1 (RM-GST) or by an FHA domain derived from the Ki67 protein. These data indicate that the FHA domain of MDC1 binds the MRE11 complex directly or indirectly, most probably in a phospho-dependent manner.

In response to DNA damage, several proteins are phosphorylated in a manner that depends on members of the phosphatidylinositol-3-OH kinase-like kinase (PIKK) family¹⁷: ATM, ATR (for ataxia-telangiectasia and Rad3-related) and/or the DNA-dependent protein kinase (DNA-PK). We found that IR exposure of cells led to reduced mobility of all forms of MDC on SDS-polyacrylamide gels (Fig. 2a). Phosphatase treatment of the cell lysate eliminated this mobility shift (Fig. 2a, top panel), indicating that IR leads to MDC1 hyperphosphorylation. Significantly, irradiation did not enhance or diminish co-immunoprecipitation of MDC1 and the MRE11 complex (data not shown). Thus, MDC1 phosphorylation must modulate some other aspect of its function. No IR-induced hyperphosphorylation of MDC1 was observed in AT cells, but phosphorylation was restored when these cells were complemented with wild-type ATM (Fig. 2a). Thus, IR-induced hyperphosphorylation of MDC1 is dependent on ATM.

To investigate further the behaviour of MDC1 in the DNA damage response, we carried out immunofluorescence studies. Whereas untreated cells showed diffuse nuclear staining of MDC1, within 15 min of exposing cells to IR or the radio-mimetic agent phleomycin, MDC1 became localized to prominent nuclear foci (Fig. 2b and data not shown). Focus formation was dose-dependent, with the average number of MDC1 foci per cell increasing with IR dosage (data not shown). MDC1 foci were still present up to 12 h later, but then gradually disappeared and were no longer evident 24 h after irradiation (data not shown). As shown in

Fig. 2b, MDC1 foci show striking colocalization with foci formed by the MRE11 complex; this occurred rapidly after irradiation (15 min) and persisted for many hours. IR-induced foci of the MRE11 complex (termed MRE11-IRIF)¹⁸ localize with foci formed by the DNA damage response protein 53BP1 and phosphorylated histone H2AX (γ -H2AX)^{19,20}, and are widely accepted to represent sites of DNA double strand breaks (DSBs)²¹. Consistent with this, MDC1-IRIF colocalized with both γ -H2AX and 53BP1 foci (Fig. 2c), but not with IR-induced foci of the homologous recombination protein RAD51 (Fig. 2c). These results show that MDC1 rapidly colocalizes with the MRE11 complex, 53BP1 and γ -H2AX at sites of IR-induced DNA DSBs.

Formation of foci by the MRE11 complex, γ -H2AX and 53BP1 depends on the activities of ATM, ATR and/or DNA-PK^{18–20,22}. Pretreatment of cells with the drug wortmannin at a concentration known to inhibit these kinases (200 μ M)²³ essentially abolished the formation of MDC1- and MRE11-IRIF (Fig. 3a and data not shown). By contrast, MDC1 foci still formed with apparently normal kinetics in cells deficient in ATM or DNA-PK, and in cells in which ATR expression was downregulated by small interfering RNA (siRNA) transfection²⁴ (Fig. 3b). Taken together, these data indicate that MDC1 focus formation can be triggered by more than one member of the PIKK family, and that MDC1-IRIF formation does not require ATM-dependent MDC1 hyperphosphorylation.

In view of the interaction between MDC1 and the MRE11 complex, we explored the possibility that these factors might rely on one another for IRIF formation. Although NBS cells failed to form MRE11-IRIF and showed predominantly cytoplasmic MRE11 staining⁶, MDC1 foci formed with apparently normal kinetics in these cells (Fig. 3c). Thus, MDC1 can be targeted to sites of DNA damage in the absence of a functional MRE11 complex.

To test whether MDC1 recruits the MRE11 complex to foci, we transfected HeLa cells with siRNA duplexes²⁴ specific for MDC1. Immunoblotting revealed that this reduced overall MDC1 levels by >90%, as compared with control cells (Fig. 5c, top panel), and immunofluorescence analysis showed that, after transfection with

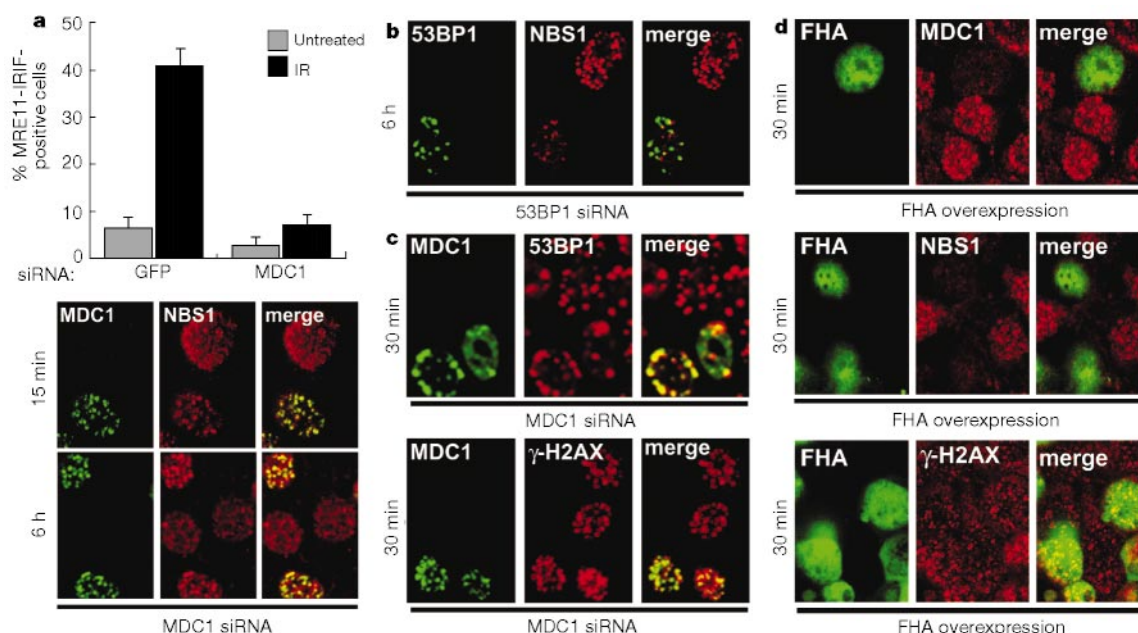


Figure 4 MDC1 is required to recruit the MRE11 complex to foci. **a**, Abrogation of MRE11-IRIF formation in MDC1-depleted cells. Top, histogram bars correspond to average numbers of NBS1-IRIF from three experiments (6 h after irradiation). Counting was carried out blind with only the NBS1-staining channel active and MDC1 siRNA data were corrected so that cells that still expressed MDC1 did not contribute to the results. Bottom, 90 h after siRNA-transfection, cells were irradiated (15 Gy) and left to recover for

15 min or 6 h before analysis. **b**, NBS1-IRIF in 53BP1-depleted cells. 53BP1-siRNA-transfected cells were analysed as above (6 h after irradiation). **c**, 53BP1 and γ -H2AX form foci in MDC1-depleted cells. MDC1-depleted cells were analysed 30 min after damage. **d**, Impaired MRE11-IRIF formation in cells overexpressing MDC1-FHA-GFP. 293T cells were transfected with MDC1-FHA-GFP and analysed 30 min after irradiation (15 Gy).

MDC1 siRNA, MDC1 was almost undetectable in most cells (Fig. 4a, c and data not shown). Next, we transfected cells with siRNA directed against MDC1, or against green fluorescent protein (GFP) as a control, irradiated the cells or left the cells untreated, and then carried out immunofluorescence studies with antibodies against MDC1 and NBS1. This revealed that cells in which MDC1 is downregulated are markedly impaired in their ability to form MRE11-IRIF, both soon (15 min) after irradiation and later on (6 h after treatment) (Fig. 4a, histogram and images). MRE11-IRIF formed normally in GFP siRNA transfected controls (Fig. 4a, histogram) and in cells in the MDC1-siRNA-transfected population that still retained MDC1 expression (Fig. 4a, images). In contrast to the above, the formation of both MDC1-IRIF and MRE11-IRIF was not affected by the siRNA-mediated downregulation of 53BP1 (Fig. 4b and data not shown), and both γ -H2AX- and 53BP1-IRIF still formed in cells in which MDC1 was downregulated (Fig. 4c). Taken together, these results show that MDC1 depletion specifically impairs MRE11-IRIF formation. Furthermore, they suggest that 53BP1 and MDC1 function in distinct branches of the DNA damage response.

Because we had found that the MDC1 FHA domain can retrieve the MRE11 complex from cell extracts, we tested the effect of overexpressing this domain in cells as a GFP fusion. Overexpression of the wild-type FHA domain markedly reduced IRIF formation by both MDC1 and the MRE11 complex (88% and 66%, respectively;

Fig. 4d), as compared with untransfected cells, whereas overexpression of the FHA domain mutant (see earlier) reduced foci formation to a lesser degree (82% for MDC1 foci and 48% for NBS1 foci). By contrast, overexpression of the MDC1 FHA domain did not impair focus formation by γ -H2AX or 53BP1 (Fig. 4d and data not shown). The overexpressed MDC1 FHA domain did not form foci on its own, indicating that this domain does not bind directly to γ -H2AX. Taken together with the protein-interaction studies, these results lead to a model in which the MDC1 FHA domain is embedded in an interaction network (both phospho-dependent and phospho-independent) that facilitates the focal concentration of the MRE11 complex at sites of DNA damage. It is noteworthy that the FHA domain of CHK2 also mediates both phospho-dependent and phospho-independent interactions²⁵.

Abrogation of MRE11-IRIF formation in AT-LD and NBS cells correlates with defects in the IR-induced S-phase checkpoint^{1,8}. Strikingly, cells in which MDC1 was downregulated by siRNA showed a clear RDS phenotype at low and high IR doses (Fig. 5a). RDS was also induced by the overexpression of the wild-type MDC1 FHA domain and, to a lesser degree, by the FHA domain mutant (Fig. 5b). These findings implicate the rapid, MDC1-mediated recruitment of the MRE11 complex to sites of DNA damage as an essential step for the efficient induction and/or propagation of the intra-S-phase checkpoint, and show that the MDC1 FHA domain has a key role in these events.

Previous work has shown that several proteins, including NBS1, CHK2 and SMC1, are phosphorylated by ATM in response to DNA damage, and that these events have important roles in the DNA damage response^{2-5,26,27,28}. Significantly, we found that although downregulation of MDC1 led to RDS, it did not impair the phosphorylation of NBS1, as detected by examining NBS1 electrophoretic mobility (Fig. 5c) or by testing for reactivity against an antibody specific for NBS1 phosphorylated on Ser343 (data not shown). These results are consistent with the finding that NBS1 phosphorylation, but not MRE11-IRIF formation, is intact in AT-LD3 cells²⁹, which display a clear RDS phenotype⁸. Notably, IR-induced phosphorylation of CHK2 on Thr68, degradation of CDC25A (Fig. 5c), and SMC1 phosphorylation on Ser966 and Ser957 (data not shown) were still observed in MDC1-depleted cells. Taken together with previous reports^{3,4,26,28}, these data reveal that CDC25A degradation and the phosphorylation of NBS1, CHK2 and SMC1 are required but not sufficient for the intra-S-phase checkpoint response.

Although previous work has suggested that MDC1 (KIAA0170/NFBD1) might be a nuclear transcriptional activator³⁰, our findings clearly demonstrate its involvement in the early stages of the DNA damage response. Specifically, MDC1 interacts with the MRE11 complex and facilitates the recruitment of this complex to foci at sites of DNA damage. We have also shown that MDC1 is required for the efficient induction of the S-phase checkpoint. MDC1 therefore joins 53BP1 and BRCA1 as a member of a novel class of checkpoint factors termed 'mediators', which may function to recruit other factors to foci at sites of DNA damage. Finally, our data indicate that two distinct and largely independent events must take place before the intra-S-phase checkpoint is engaged. One corresponds to MDC1-mediated recruitment of the MRE11 complex and, possibly, other factors to IRIF. The other consists of the activation of the ATM pathway and the ensuing phosphorylation of downstream proteins^{2-5,26,27,28}. We speculate that this dual triggering mechanism might have evolved to endow the DNA damage response with the necessary high degree of biological specificity, and to prevent this pathway from being fully launched inappropriately. □

Methods

Identification of MDC1

Sheep anti-MRE11 and anti-RAD50 antisera were raised against purified (His)₆-tagged fragments of the proteins (residues 183–583 of MRE11 and 480–724 of RAD50) and

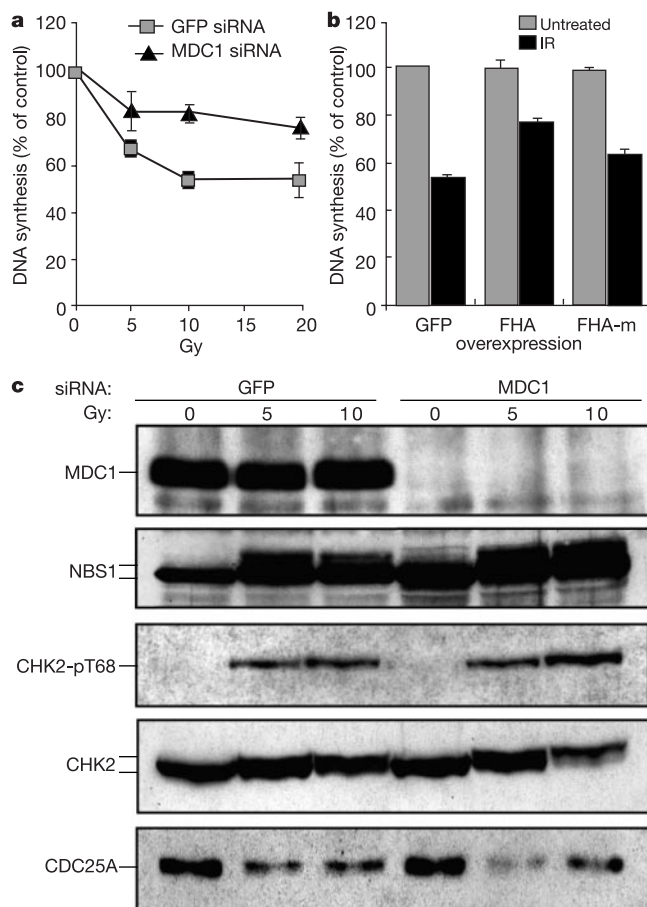


Figure 5 MDC1 functions in the S-phase checkpoint. **a**, RDS phenotype in cells in which MDC1 was downregulated. siRNA-transfected cells were assessed for DNA synthesis 45 min after irradiation. **b**, RDS phenotype in cells overexpressing MDC1-FHA-GFP. HeLa cells were transfected with MDC1-FHA-GFP, MDC1-FHA-m-GFP or GFP plasmids, irradiated with 10 Gy and analysed as above. **c**, Phosphorylation of ATM targets in cells in which MDC1 was downregulated. Cells were irradiated with the indicated doses 96 h after siRNA transfection, and left to recover for 30 min before cell extraction. Proteins were analysed by immunoblotting with the indicated antibodies.

affinity-purified on a column prepared from the same fragments coupled to SulfoLink resin (Pierce). HeLa cell nuclear extracts (Computer Cell Culture Centre) were incubated with cyanogen-bromide-activated Sepharose 4B (Amersham Pharmacia Biotech) linked to affinity-purified anti-MRE11 or anti-RAD50 antibodies. Bound proteins were separated by SDS–polyacrylamide gel electrophoresis and stained with silver. For identification of proteins, a gel was stained with Coomassie blue and the p250 protein band was excised, destained and digested in-gel with trypsin. Extracted peptides were analysed by matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (Time-of-Flight mass spectrometer, Micromass) and several peptides sequenced by low-energy collision-induced dissociation (CID) using an LCQ-Deca Ion-Trap mass spectrometer (Finnigan Corp) fitted with a nano-electrospray ion source. Peptide-mass fingerprint and peptide CID spectra were screened against the National Center for Biotechnology Information non-redundant protein sequence database using the MASCOT search engine.

Plasmids

MDC1-FHA–GST and MDC1-FHA–GFP correspond to MDC1 residues 2–220 fused to GST and GFP, respectively; MDC1-FHA–m–GST and MDC1-FHA–m–GFP contain MDC1 residues 2–220 (Arg58Ala); MDC1-BRCT–GST contains MDC1 residues 1883–2089; MDC1-RM–GST contains MDC1 residues 1459–1500; Ki67-FHA–GST contains Ki67 residues 1–134.

Antibodies, immunoprecipitation and GST pull-downs

The antibodies used in this study were: mouse and rabbit anti- γ -H2AX (Upstate Biotechnology); rabbit anti-RAD51 (Ab-1; Oncogene Research Products); rabbit p95NBS1 (Ab-1; Oncogene Research Products); rabbit anti-NBS1 (Novus Biologicals); goat anti-ATR (FRP1; Santa Cruz Biotechnology); rabbit anti-phospho-CHK2 (Thr68; Cell Signaling Technology); mouse anti-CHK2 (DCS-273); mouse anti-CDC25A (F6; Santa Cruz Biotechnology); mouse anti-PARP (Serotec); rabbit anti-SMC1, anti-SMC1 (pSer966) and anti-SMC1 (pSer957) (Bethyl Laboratories); mouse anti-53BP1 (a gift from T. D. Halazonetis); mouse anti-NBS1-S343(P) (a gift from M. Kastan); and rabbit anti-MRE11 and anti-NBS1 (gifts from J. Petrini). Polyclonal anti-MDC1(889) and anti-MDC1(3835) antibodies were raised in rabbit and sheep, respectively, against MDC1-FHA–GST. Polyclonal anti-MDC1(3838) antibodies were raised in sheep against MDC1-BRCT–GST. Anti-NBS1 polyclonal rabbit serum was raised against a purified (His)₆-tagged NBS1 fragment (residues 200–351).

GST pull-downs were done with bacterially expressed and purified MDC1 fragments and with HeLa nuclear extracts. For co-immunoprecipitation, HeLa nuclear extracts (1–2 mg), prepared from undamaged cells and from cells treated with 10 Gy of IR (Faxitron Cabinet X-ray system, model 43855D; Faxitron X-Ray Corporation), were incubated with primary antibodies in the presence of 10 μ g ml^{−1} ethidium bromide for 2 h at 4 °C. Immuno-complexes were precipitated with protein-G Sepharose (Amersham Pharmacia Biotech). Nuclear extracts (100 μ g) were incubated at 30 °C with 400 units of λ phosphatase (New England BioLabs) in the presence of Mn²⁺ for 30 min before immunoblot analyses. In controls, 50 mM EDTA was added to inhibit the phosphatase.

Cell lines and transfections

HeLa, MRC5, 293T, NBS-1LBI, M059J (DNA-PK^{−/−}) and M059K (DNA-PK^{+/+}) cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). ATT21JE-T FT-pEB57 (empty vector) and ATT21JET FT-pEBYZ5 (+ATM) cells were cultured in DMEM supplemented with 10% FBS and 100 μ g ml^{−1} hygromycin B (Invitrogen Life Technologies).

siRNA duplexes were synthesized by Dharmacon. The coding strands for MDC1, 53BP1, ATR and GFP siRNAs were AATCCTGAGACCTCCTAAGGTTT, AAGCCAGGT TCTAGAGGATGATT, AAGACGGTGTGCTCATGCCGC and AACACTTGTCACTACTT TCTC, respectively. Transfection of HeLa cells with siRNA duplexes was as described²⁴. Transfection of 293T cells with GFP-fused constructs was with Lipofectamine 2000 (Invitrogen).

Immunofluorescence microscopy

Cells were grown on glass coverslips coated with poly-L-lysine (Sigma). Fixation was either with 2% paraformaldehyde (w/v) in PBS for 10 min, after which cells were permeabilized with 0.5% NP-40 for 10 min, or with ice-cold methanol for 15 min on ice followed by permeabilization in ice-cold acetone for 20 s. Coverslips were washed with PBS and blocked in 10% FBS. After incubation with primary antibody in 5% FBS, cells were washed with PBS. For single antibody staining, fluorescein isothiocyanate (FITC)-conjugated antibodies (Jackson Laboratories) were added. Coverslips were washed with PBS and mounted with Vectashield mounting medium containing propidium iodide (Vector Labs). For dual antibody staining, FITC-conjugated antibodies and rhodamine-conjugated antibodies (Jackson Laboratories) were used. Coverslips were washed with PBS and mounted with Vectashield mounting medium (Vector Labs). Slides were viewed with a Bio-Rad confocal laser microscope by sequential scanning of the two emission channels used (488 nm for FITC, and 514 nm for propidium iodide and rhodamine).

RDS assay

48 h after siRNA-transfection, HeLa cells were labelled for 24 h with 20 nCi ml^{−1} [¹⁴C]thymidine, followed by another 24 h incubation in non-radioactive medium. Cells were irradiated with 0, 5, 10 or 20 Gy; 45 min later, [³H]thymidine (2.5 μ Ci ml^{−1}) was added. For FHA domain overexpression studies, HeLa cells were transfected using the Lipofectin and PLUS reagents as recommended by the manufacturer (Invitrogen). [¹⁴C]thymidine (20 nCi ml^{−1}) was added at the time of transfection, washed away 24 h

later, and fresh medium was added. After another 24 h, the cells were irradiated (10 Gy). After a further 45 min, [³H]thymidine (2.5 μ Ci ml^{−1}) was added for 15 min. Inhibition of DNA synthesis was then measured as described¹³.

Received 12 December 2002; accepted 24 January 2003; doi:10.1038/nature01445.

1. D'Amours, D. & Jackson, S. P. The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature Rev. Mol. Cell Biol.* **3**, 317–327 (2002).
2. Gatei, M. *et al.* ATM-dependent phosphorylation of nibrin in response to radiation exposure. *Nature Genet.* **25**, 115–119 (2000).
3. Lim, D. S. *et al.* ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway. *Nature* **404**, 613–617 (2000).
4. Wu, X. *et al.* ATM phosphorylation of Nijmegen breakage syndrome protein is required in a DNA damage response. *Nature* **405**, 477–482 (2000).
5. Zhao, S. *et al.* Functional link between ataxia–telangiectasia and Nijmegen breakage syndrome gene products. *Nature* **405**, 473–477 (2000).
6. Carney, J. P. *et al.* The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: Linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93**, 477–486 (1998).
7. Matsuura, S. *et al.* Positional cloning of the gene for Nijmegen breakage syndrome. *Nature Genet.* **19**, 179–181 (1998).
8. Stewart, G. S. *et al.* The DNA double-strand break repair gene *hMRE11* is mutated in individuals with an ataxia–telangiectasia-like disorder. *Cell* **99**, 577–587 (1999).
9. Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753 (1995).
10. Varon, R. *et al.* Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**, 467–476 (1998).
11. Michelson, R. J. & Weinert, T. Closing the gaps among a web of DNA repair disorders. *BioEssays* **22**, 966–969 (2000).
12. Shiloh, Y. Ataxia–telangiectasia and the Nijmegen breakage syndrome: Related disorders but genes apart. *Annu. Rev. Genet.* **31**, 635–662 (1997).
13. Falck, J., Petrini, J. H., Williams, B. R., Lukas, J. & Bartek, J. The DNA damage-dependent intra-S phase checkpoint is regulated by parallel pathways. *Nature Genet.* **30**, 290–294 (2002).
14. Nagase, T., Seki, N., Ishikawa, K., Tanaka, A. & Nomura, N. Prediction of the coding sequences of unidentified human genes. V. The coding sequences of 40 new genes (K1AA0161–K1AA0200) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res.* **3**, 17–24 (1996).
15. Durocher, D. & Jackson, S. P. The FHA domain. *FEBS Lett.* **513**, 58–66 (2002).
16. Bork, P. *et al.* A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.* **11**, 68–76 (1997).
17. Abraham, R. T. Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes Dev.* **15**, 2177–2196 (2001).
18. Maser, R. S., Monsen, K. J., Nelms, B. E. & Petrini, J. H. hMre11 and hRad50 nuclear foci are induced during the normal cellular response to DNA double-strand breaks. *Mol. Cell. Biol.* **17**, 6087–6096 (1997).
19. Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *J. Cell Biol.* **151**, 1381–1390 (2000).
20. Paull, T. T. *et al.* A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10**, 886–895 (2000).
21. Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. *J. Cell Biol.* **146**, 905–916 (1999).
22. Burma, S., Chen, B. P., Murphy, M., Kurimasa, A. & Chen, D. J. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J. Biol. Chem.* **276**, 42462–42467 (2001).
23. Sarkaria, J. N. *et al.* Inhibition of phosphoinositide 3-kinase related kinases by the radiosensitizing agent wortmannin. *Cancer Res.* **58**, 4375–4382 (1998).
24. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
25. Li, J. *et al.* Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumour suppressor kinase Chk2. *Mol. Cell* **9**, 1045–1054 (2002).
26. Falck, J., Mailand, N., Syljuasen, R. G., Bartek, J. & Lukas, J. The ATM–Chk2–Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature* **410**, 842–847 (2001).
27. Yazdi, P. T. *et al.* SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev.* **16**, 571–582 (2002).
28. Kim, S. T., Xu, B. & Kastan, M. B. Involvement of the cohesin protein, SMC1, in Atm-dependent and independent responses to DNA damage. *Genes Dev.* **16**, 560–570 (2002).
29. Maser, R. S. *et al.* Mre11 complex and DNA replication: Linkage to E2F and sites of DNA synthesis. *Mol. Cell Biol.* **21**, 6006–6016 (2001).
30. Ozaki, T. *et al.* NFB1/KIAA0170 is a novel nuclear transcriptional transactivator with BRCT domain. *DNA Cell Biol.* **19**, 475–485 (2000).

Acknowledgements We thank Y. Shiloh for AT complemented cells, T. D. Halazonetis for anti-53BP1 antibodies, J. Petrini for anti-MRE11 and anti-NBS1 antibodies, and M. Kastan and K. K. Khanna for anti-NBS1-S343(P) antibodies. We thank S.P.J. laboratory members for helpful discussions and critical reading of the manuscript, especially J. Bradbury, F. d'Adda di Fagagna, M. Grenon, J. Rouse and V. Smits. We thank S. Elledge for sharing unpublished results. This study was supported by grants from the Association for International Cancer Research, the A-T Medical Research Trust and the Swiss National Foundation. Research in the S.P.J. laboratory is funded by Cancer Research UK.

Competing interests statement The authors declare competing financial interests: details accompany the paper on *Nature's* website (► <http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to S.P.J. (e-mail: spj13@mole.bio.cam.ac.uk).

MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways

Zhenkun Lou, Katherine Minter-Dykhouse, Xianglin Wu & Junjie Chen

Department of Oncology, Mayo Foundation, Rochester, Minnesota 55905, USA

Forkhead-homology-associated (FHA) domains function as protein–protein modules that recognize phosphorylated serine/threonine motifs^{1–5}. Interactions between FHA domains and phosphorylated proteins are thought to have essential roles in the transduction of DNA damage signals; however, it is unclear how FHA-domain-containing proteins participate in mammalian DNA damage responses. Here we report that a FHA-domain-containing protein—mediator of DNA damage checkpoint protein 1 (MDC1; previously known as KIAA0170)—is involved in DNA damage responses. MDC1 localizes to sites of DNA breaks and associates with CHK2 after DNA damage. This association is mediated by the MDC1 FHA domain and the phosphorylated Thr 68 of CHK2. Furthermore, MDC1 is phosphorylated in an ATM/CHK2-dependent manner after DNA damage, suggesting that MDC1 may function in the ATM–CHK2 pathway. Consistent with this hypothesis, suppression of MDC1 expression results in defective S-phase checkpoint and reduced apoptosis in response to DNA damage, which can be restored by the expression of wild-type MDC1 but not MDC1 with a deleted FHA domain. Suppression of MDC1 expression results in decreased p53 stabilization in response to DNA damage. These results suggest that MDC1 is recruited through its FHA domain to the activated CHK2, and has a critical role in CHK2-mediated DNA damage responses.

The DNA-damage-induced cell-cycle checkpoints and apoptosis are critical for the maintenance of genomic stability^{6,7}. Many proteins involved in DNA damage responses contain functional domains such as FHA and BRCA1 carboxy-terminal (BRCT) domains^{1,8}. FHA and BRCT domains are molecular modules involved in protein–protein interaction. FHA domains specifically recognize motifs that contain phosphorylated serine/threonine in a manner similar to the SH2-phosphotyrosine interaction^{2–5}. A well-studied example of the FHA-domain-mediated interaction is the Rad9–Rad53 interaction in *Saccharomyces cerevisiae*^{9–11}, which is critical for the activation of DNA damage pathways in yeast. However, except for the homo-oligomerization of checkpoint kinase 2 (CHK2)^{12,13}, such FHA-domain-mediated interactions are not documented in mammalian cells. More importantly, the role of FHA domains in DNA damage signal transduction remains to be established in mammals.

MDC1 is a protein that contains both FHA and BRCT domains. Because the BRCT domain of MDC1 shares close sequence homology with that of BRCA1 and yeast Rad9, we examined whether MDC1 is involved in DNA damage response pathways in mammalian cells. Many proteins involved in DNA damage response pathways form nuclear foci after DNA damage (for example, BRCA1 and CHK2)^{14,15}. Therefore, we first investigated the nuclear localization of MDC1 before and after DNA damage. MDC1 normally localized in the nuclei without any apparent structures. However, after γ -irradiation, MDC1 relocalized to nuclear foci (Fig. 1a) and localized together with phosphorylated H2AX (γ H2AX) foci, suggesting that MDC1 relocalizes to the sites of DNA damage^{16,17}. In addition, similar to many other signalling proteins in DNA damage response pathways, MDC1 migrates slower after γ -irradiation (Fig. 1b). Treatment of MDC1 immunoprecipitates with λ -phosphatase reverses this mobility shift (Fig. 1b), suggesting

that MDC1 is phosphorylated after γ -irradiation. To determine whether the phosphorylation of MDC1 depends on ataxia–telangiectasia mutated (ATM), K562 cells and ATM-deficient GM03189D cells were either irradiated or left untreated. In contrast to that in K562 cells, no mobility shift of MDC1 was observed in GM03189D cells after γ -irradiation (Fig. 1c), suggesting that the phosphorylation of MDC1 is ATM-dependent.

To explore the role of MDC1 in DNA damage pathways, we examined whether MDC1 would associate with proteins involved in DNA damage response pathways. As shown in Fig. 2a, MDC1 associated with CHK2 only after DNA damage. This interaction is not interrupted by ethidium bromide (data not shown), suggesting that the MDC1–CHK2 interaction is not mediated by DNA. Furthermore, MDC1 localized together with the phosphorylated Thr 68 residue of CHK2 after DNA damage (Fig. 2b). As MDC1 interacts with the active form of CHK2 (Thr-68-phosphorylated CHK2) after DNA damage, we investigated whether the phosphorylation of MDC1 requires CHK2. CHK2-deficient HCT15 cells and HCT15 cells stably transfected with either wild-type CHK2 or kinase-dead CHK2 were irradiated or left untreated. No mobility shift of MDC1 was observed in HCT15 cells after γ -irradiation. However, mobility shift of MDC1 was restored in HCT15 cells reconstituted with wild-type CHK2, but not in cells reconstituted with kinase-dead CHK2 (Fig. 2c).

To explore further the interaction between MDC1 and CHK2, we generated glutathione S-transferase (GST) fusion proteins that contained either the MDC1 FHA domain or the MDC1 BRCT domain, as FHA and BRCT domains are involved in protein–protein interactions. Figure 2d shows that GST–MDC1(FHA), but not GST–MDC1(BRCT), interacted with CHK2 after DNA damage. The *in vitro* interactions of GST–MDC1(FHA) and CHK2 are

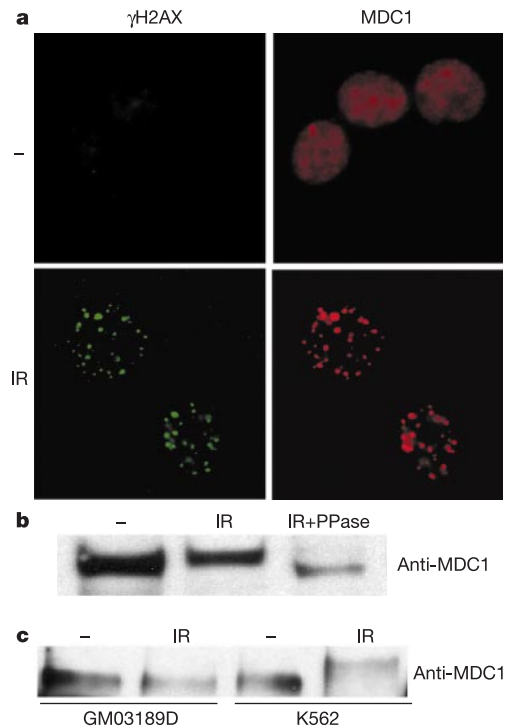


Figure 1 MDC1 is phosphorylated and relocalizes to nuclear foci after DNA damage. **a**, 293T cells were left untreated (top panels) or treated with 10 Gy γ -irradiation (IR; bottom panels), and 10 h later they were stained with anti- γ H2AX antibodies (left) or anti-MDC1 antibodies (right). **b**, MDC1 was immunoprecipitated from untreated or γ -irradiated (30 Gy) K562 cells and incubated with buffer or λ -phosphatase (PPase), then blotted with anti-MDC1 antibodies. **c**, K562 cells and GM03189D cells (ATM deficient) were γ -irradiated (30 Gy) or left untreated, then blotted with anti-MDC1 antibodies.

specific and depend on the integrity of the FHA domain of MDC1, as mutations in highly conserved residues within this domain (R58A, S72A, N96G97T98/AAA) abolished or reduced the interaction between MDC1 and CHK2 (data not shown). To test whether the MDC1 FHA domain is required for the MDC1–CHK2 interaction *in vivo*, we generated constructs encoding S-tagged wild-type MDC1 or MDC1 with a deleted FHA domain (MDC1 Δ FHA). 293T cells transfected with these constructs were irradiated and lysates were incubated with S-Sepharose to pull-down wild-type or the deletion mutant of MDC1. As shown in Fig. 2e, in contrast with wild-type MDC1, MDC1 Δ FHA failed to interact with CHK2. These results suggest that the damage-dependent interaction between MDC1 and CHK2 depends on the FHA domain of MDC1.

To examine which region of CHK2 is important for the MDC1–CHK2 interaction, constructs encoding haemagglutinin (HA)-tagged wild-type CHK2 (CHK2WT) and CHK2 deleted of the S/TQ-rich region (CHK2 Δ S/TQ) were stably transfected into CHK2-deficient HCT15 cells. Deletion of the S/TQ-rich region of CHK2 completely abolished the MDC1–CHK2 association, suggesting that the S/TQ-rich region of CHK2 is required for MDC1 binding (data not shown). The S/TQ-rich region of CHK2 contains multiple potential phosphorylation sites for ATM/ataxia–telangiectasia and Rad3 related (ATR) kinases. Among them, Thr 68 has been shown to be an ATM phosphorylation site that is critical for CHK2 activation^{18–22}. Notably, as determined by degenerate peptide library studies, the CHK2 sequence surrounding phosphorylated Thr 68 is a predicated recognition sequence of the MDC1 FHA domain⁴. Therefore, we next tested whether phosphorylated Thr 68 of CHK2

is the site recognized by the MDC1 FHA domain. Again, *in vitro* pull-down assay using GST–MDC1(FHA) was performed using HCT15 cells stably transfected with CHK2WT, kinase-dead CHK2 (CHK2KD) or CHK2 containing a T68A mutation (CHK2T68A). Whereas the kinase-dead mutant of CHK2 still interacted with MDC1 in a DNA-damage-dependent manner (Fig. 2f), the T68A mutation abolished the MDC1–CHK2 interaction. These results suggest that the phosphorylation of CHK2 at Thr 68 by ATM, but not CHK2 autophosphorylation sites, is critical for MDC1 binding.

To confirm further the critical role of Thr-68-phosphorylated CHK2 in the MDC1–CHK2 interaction, synthetic CHK2 peptides representing CHK2 amino acid sequence surrounding Thr 68 (residues 64–73) were prepared with Thr 68 as either a phosphorylated (p-T68) or non-phosphorylated (T68) residue. If phosphorylated Thr 68 and the surrounding sequence of CHK2 are truly recognized by the MDC1 FHA domain, there would be competition between the p-T68 peptide and endogenous CHK2 for binding to GST–MDC1(FHA). As expected, p-T68 peptide, but not T68 peptide, effectively blocked the binding of endogenous CHK2 to GST–MDC1(FHA) after DNA damage (Fig. 2g). As a negative control, a phosphorylated peptide containing CHK2 Ser 516 (an autophosphorylation site; our own unpublished observation) did not block the binding of CHK2 with MDC1 (Fig. 2g). Furthermore, when p-T68 or T68 peptides were conjugated to Sepharose and incubated with cell lysates, only p-T68 peptide was able to interact with MDC1 (Fig. 2h). To demonstrate the direct interaction between the MDC1 FHA domain and the phosphorylated Thr 68 site of CHK2, T68 or p-T68 peptide conjugated with Sepharose were

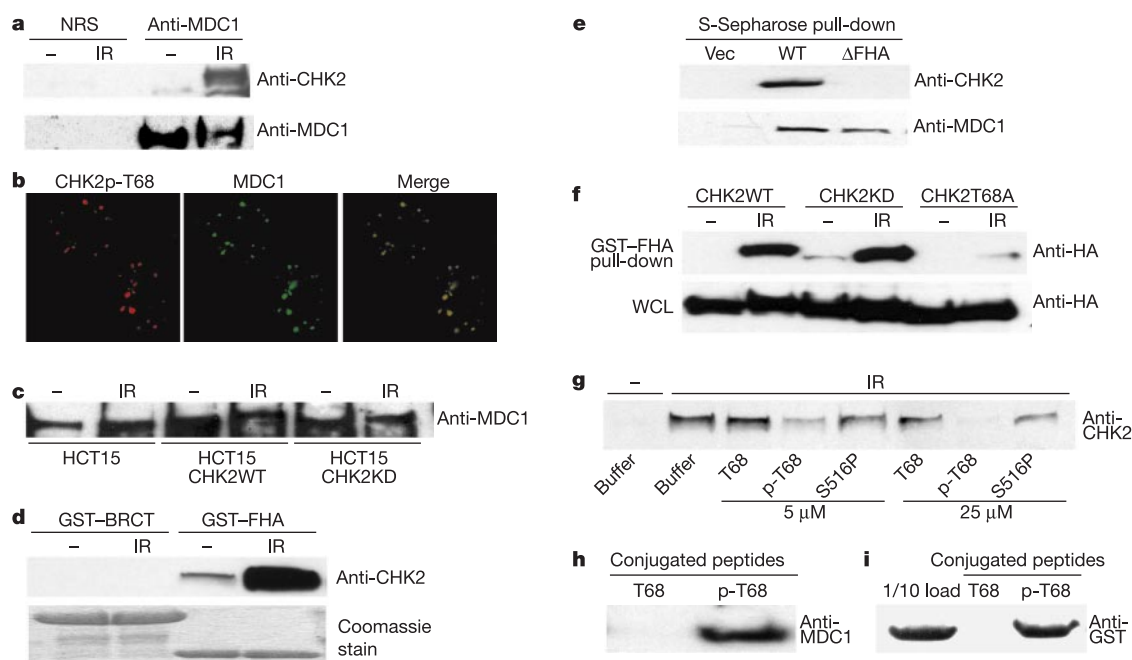


Figure 2 MDC1 interacts with CHK2 through the MDC1 FHA domain and the phosphorylated Thr 68 of CHK2. **a**, Cell lysates from K562 cells untreated or treated with 30 Gy γ -irradiation were subjected to immunoprecipitation with normal rabbit serum (NRS) or anti-MDC1 antibodies, and the immunoprecipitates were analysed with anti-CHK2 antibody. **b**, 293T cells were irradiated (10 Gy) and 10 h later stained with the indicated antibodies. **c**, HCT15 cells (CHK2 deficient) and HCT15 cells stably transfected with wild-type CHK2 or kinase-dead CHK2 were γ -irradiated (30 Gy) or left untreated, and cell lysates were then blotted with anti-MDC1 antibodies. **d**, *In vitro* binding assay. Beads bound with GST–MDC1(FHA) and GST–MDC1(BRCT) were incubated with lysates from control or γ -irradiated (30 Gy) K562 cells. Binding of CHK2 was analysed by immunoblotting. **e**, 293T cells were transfected with the indicated constructs and treated with 30 Gy γ -irradiation. The association of S-tagged MDC1 with CHK2 was assessed

using the indicated antibodies. WT, wild type. **f**, *In vitro* binding assays. HCT15 cells stably transfected with the indicated HA-tagged constructs were treated with 30 Gy γ -irradiation or left untreated. Beads bound with GST–MDC1(FHA) were used to pull-down HA-tagged wild-type or mutants, and were blotted with anti-HA antibodies. Whole cell lysates (WCL) were blotted with anti-HA antibodies (bottom panel). **g**, Peptide competition assay. K562 cell lysates were incubated with beads bound with GST–MDC1(FHA) plus the indicated concentration of CHK2 peptides. The remaining CHK2 on the beads was detected by anti-CHK2 antibodies. **h**, **i**, Non-phosphorylated or phosphorylated Thr peptides (T68 and p-T68, respectively) conjugated to Sepharose beads were incubated with the lysate prepared from untreated K562 cells (**h**), or purified GST–MDC1(FHA) (**i**). Proteins bound on beads were analysed with anti-MDC1 antibodies (**h**) or anti-GST antibodies (**i**).

incubated with purified GST-MDC1(FHA). As shown in Fig. 2i, only p-T68 peptide strongly binds to the MDC1 FHA domain. These data suggest that phosphorylated Thr 68 of CHK2 acts as a docking site for MDC1.

The interaction of MDC1 and CHK2 after γ -irradiation suggests

that MDC1 may be involved in CHK2-mediated DNA damage responses. To investigate whether MDC1 functions in the ATM-CHK2 pathway, we used the small interfering RNA (siRNA) technique to suppress MDC1 expression²³. Transfection of MDC1 siRNA resulted in the suppression of MDC1 in A549 cells (Fig. 3a, b).

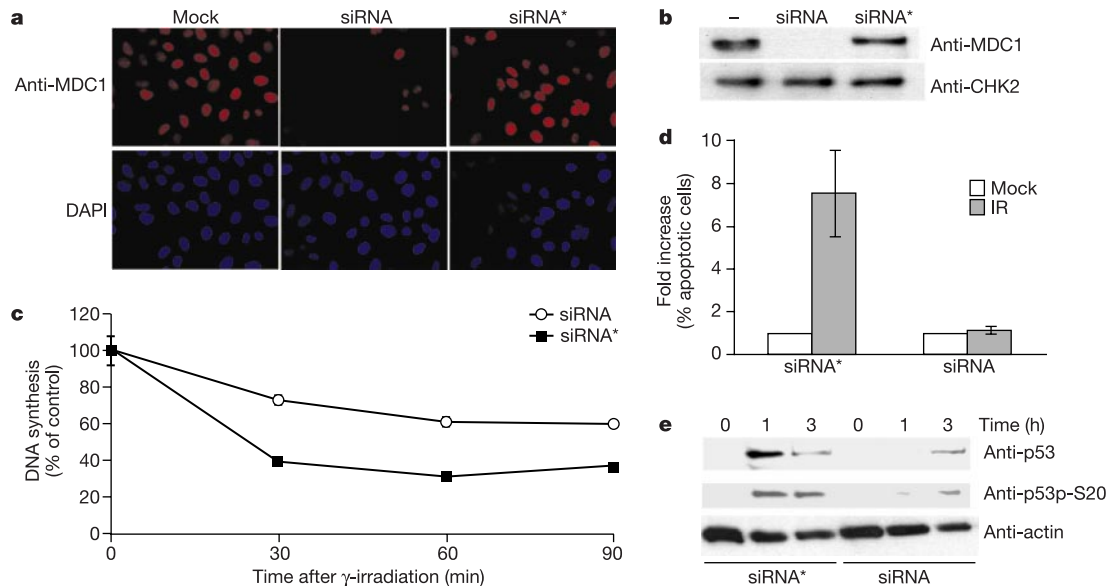


Figure 3 MDC1 is required for CHK2-mediated DNA damage responses. **a, b**, A549 cells were mock-transfected or transfected with MDC1 siRNA or control siRNA (siRNA*) twice. Seventy-two hours after the initial transfection, cells were immunostained with anti-MDC1 antibodies (**a**) or blotted with anti-MDC1 antibodies or CHK2 antibodies as a control (**b**). DAPI, 4,6-diamidino-2-phenylindole. **c**, A549 cells transfected with MDC1 siRNA or control siRNA as described for **a** and **b**. Cells were γ -irradiated (20 Gy) or left untreated,

and DNA synthesis was assessed. **d**, A549 cells were transfected with either control siRNA or MDC1 siRNA, then left untreated or irradiated (5 Gy). Thirty-six hours later, cells were collected and evaluated for apoptosis. **e**, A549 cells were transfected with either control siRNA or MDC1 siRNA, then irradiated (5 Gy). Cells were lysed at the time indicated and blotted for p53, Ser-20-phosphorylated (p-S20) p53 or actin.

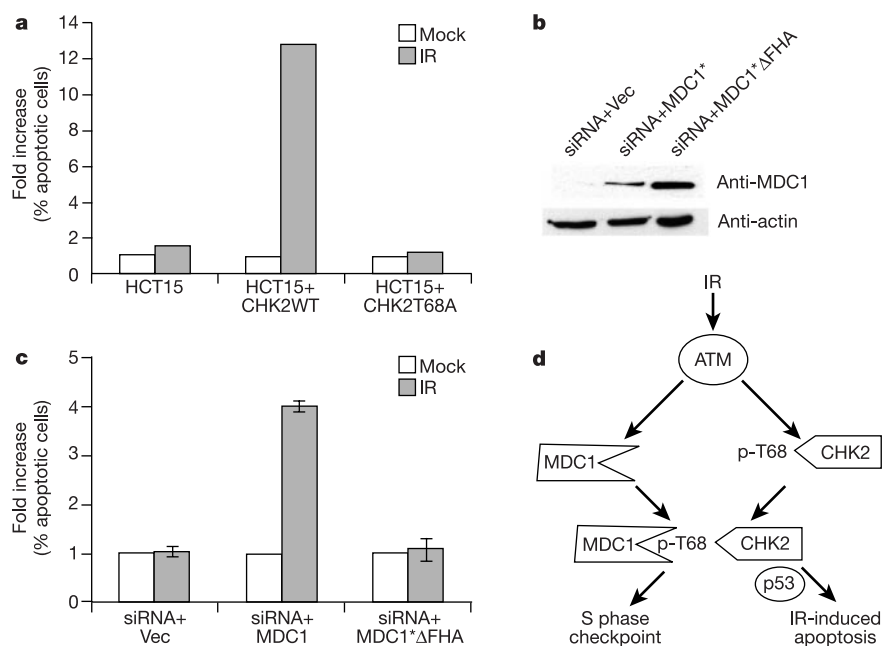


Figure 4 The functional interaction of CHK2 and MDC1 is required for CHK2-mediated DNA damage responses. **a**, HCT15 cells (CHK2 deficient) and HCT15 cells stably transfected with wild-type CHK2 or CHK2T68A were γ -irradiated (10 Gy) or left untreated. Thirty-six hours later, cells were collected and evaluated for apoptosis. **b, c**, A549 cells were transfected with MDC1 siRNA, and 24 h later they were transfected with constructs encoding green fluorescent protein (GFP) and either pCDNA3 vector, pCDNA3 encoding

siRNA-resistant wild-type MDC1 (MDC1*) or MDC1 Δ FHA (MDC1* Δ FHA). Thirty-six hours after the initial transfection, cells were collected and blotted for MDC1 expression (**b**), or sorted for GFP-positive cells (**c**). The GFP-positive cells were then either irradiated (5 Gy) or left untreated, and 48 h later, cells were evaluated for apoptosis. **d**, Model of how MDC1 interacts with activated CHK2 and functions in CHK2-mediated DNA damage responses.

Similar effects were observed in HeLa cells, BJ cells and 293 T cells (data not shown). We also prepared control siRNA by introducing four point mutations into MDC1 siRNA (siRNA*). Transfection of control siRNA did not affect the level of MDC1, indicating the specificity of MDC1 siRNA (Fig. 3a, b). To examine whether suppression of MDC1 expression affects CHK2-mediated DNA damage responses, we examined downstream events regulated by CHK2. One such event is the ATM/CHK2-dependent temporal inhibition of DNA synthesis in S phase^{24,25}. In ATM-deficient and CHK2-deficient cells, radioresistant DNA synthesis (RDS) occurs owing to the defective S-phase checkpoint^{24,25}. We examined DNA synthesis after γ -irradiation in cells transfected with MDC1 siRNA or control siRNA. There was a 60–70% inhibition of DNA synthesis in cells transfected with control siRNA after γ -irradiation (Fig. 3c). However, in cells transfected with MDC1 siRNA, there was only 30–40% inhibition, consistent with an RDS phenotype (Fig. 3c). Similar RDS phenotype was also observed in BJ cells transfected with MDC1 siRNA (data not shown). To evaluate further the role of MDC1 in CHK2-mediated DNA responses, we evaluated the radiation-induced apoptosis. Recent studies suggest that the radiation-induced apoptosis requires CHK2 activity^{26–28}. As shown in Fig. 3d, γ -irradiation induced apoptosis in A549 cells transfected with control siRNA. This γ -irradiation-induced apoptosis was abolished when A549 cells were transfected with MDC1 siRNA. These results indicate that MDC1 has a critical involvement in CHK2-mediated DNA damage responses. To explore further downstream molecules regulated by MDC1, we examined p53 stabilization after γ -irradiation, as CHK2 regulates radiation-induced apoptosis by phosphorylating p53 at Ser 20, which leads to the stabilization of p53 (refs 26–30). Compared with cells transfected with control siRNA, cells transfected with MDC1 siRNA show decreased and delayed p53 stabilization after γ -irradiation (Fig. 3e). Consistent with p53 stabilization, phosphorylation of p53 at Ser 20 is also decreased and delayed in cells transfected with MDC1 siRNA (Fig. 3e). These results indicate that MDC1 was able to regulate γ -irradiation-induced apoptosis by influencing p53 stabilization.

To establish specifically that the CHK2–MDC1 interaction is required for the DNA damage responses shown above, we evaluated γ -irradiation-induced apoptosis in HCT15 cells stably transfected with wild-type CHK2 or CHK2 with a mutation at Thr 68 (T68A), which could not associate with MDC1. As shown in Fig. 4a, HCT15 cells show defective apoptosis in response to γ -irradiation, which was restored by the reconstitution of wild-type CHK2, but not CHK2T68A. Furthermore, we generated MDC1 constructs (full-length and Δ FHA) that are resistant to siRNA targeting by introducing silent mutations at the region targeted by siRNA. Co-transfection of these constructs with MDC1 siRNA can restore the expression of MDC1 (Fig. 4b). Co-transfection of MDC1 siRNA and siRNA-resistant full-length MDC1 (MDC1*), but not siRNA-resistant MDC1 Δ FHA (MDC1* Δ FHA), restored γ -irradiation-induced apoptosis (Fig. 4c). These results suggest that the interaction of CHK2 and MDC1 is critical for CHK2-mediated DNA damage responses.

We have shown that MDC1 is recruited to activated CHK2 in response to DNA damage. This recruitment is conducted through the interaction between the FHA domain of MDC1 and the phosphorylated Thr 68 site of CHK2. MDC1 is phosphorylated in an ATM- and CHK2-dependent manner and is involved in S-phase checkpoint and radiation-induced apoptosis (Fig. 4d). These results indicate that MDC1 has an important role in CHK2-mediated DNA damage responses. Furthermore, these data demonstrate the importance of the FHA domain–phosphoprotein interaction in the mammalian DNA damage signalling pathways. □

Methods

Plasmids and siRNAs

MDC1 complementary DNA was provided by T. Nagase (Kazusa DNA Research Institute). Wild-type MDC1 or MDC1 Δ FHA was cloned into a pIRES2 vector containing S-protein tag. siRNA-resistant constructs were made by introducing a silent mutation at the MDC1 coding region (70–75; TTGAGG to CTGAGA). The MDC1 FHA and BRCT domains were cloned into the pGEX5.3 vector for bacterial expression of GST fusion proteins. CHK2 constructs have been described previously¹⁵. MDC1 siRNAs were synthesized by Xerogen Inc. The siRNA duplexes were 21 base pairs including a 2-deoxynucleotide overhang. The coding strand of MDC1 siRNA (MDC1 cDNA 58–76) was UCCAGUGAAUCCUUGAGGUdTdT; the control siRNA was UUCAUAAAUCUUGAGGUdTdT.

Antibodies and cell lines

MDC1 antibodies were raised against GST fusion proteins containing amino-terminal residues 1–150 or 151–484 of MDC1. CHK2 antibodies have been previously described¹⁵. Anti- γ H2AX antibodies were generated as previously described¹⁵. Antibodies against p53 and phosphorylated Ser 20 of p53 were purchased from Cell Signaling.

All cells were obtained from American Tissue Culture Collections and cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum. HCT15 stable cell lines expressing wild-type or mutant CHK2 were generated by neomycin selection.

Immunoprecipitation, immunoblotting and immunostaining

We performed cell lysate preparation, immunoprecipitation, immunoblotting and immunostaining as described¹⁴.

In vitro pull-down assay and peptide competition assay

GST fusion proteins were bound to glutathione-Sepharose overnight at 4 °C. The beads were washed with PBS twice and incubated with cell lysates for 1 h at 4 °C. For peptide competition assay, the indicated concentration of T68 peptide (CETVSTQELYS) and p-T68 peptide (CETVS-pT-QELYS) was added to cell lysates before incubation with beads bound to GST fusion proteins.

siRNA transfection, RDS assay and apoptosis assay

siRNA transfection was performed as described previously²³. Seventy-two hours after initial transfection, cells were γ -irradiated (20 Gy) or left untreated. Cells were then pulsed with [³H]thymidine after the indicated times, and collected 30 min later. For apoptosis assay, cells were irradiated at the indicated doses, and 36 h later fixed and stained with Hoechst. Cells were then scored for apoptosis.

Received 12 December 2002; accepted 27 January 2003; doi:10.1038/nature01447.

- Hofmann, K. & Bucher, P. The FHA domain: a putative nuclear signaling domain found in protein kinases and transcription factors. *Trends Biochem. Sci.* **20**, 347–349 (1995).
- Durocher, D., Henkel, J., Fersht, A. R. & Jackson, S. P. The FHA domain is a modular phosphopeptide recognition motif. *Mol. Cell* **4**, 387–394 (1999).
- Wang, P. *et al.* Structure and specificity of the interaction between the FHA2 domain of Rad53 and phosphorylated peptides. *J. Mol. Biol.* **302**, 927–940 (1999).
- Durocher, D. *et al.* The molecular basis of FHA domain: phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell* **6**, 1169–1182 (2000).
- Li, J., Smith, G. P. & Walker, J. C. Kinase interaction domain of kinase-associated protein phosphatase, a phosphoprotein-binding domain. *Proc. Natl Acad. Sci. USA* **96**, 7821–7826 (1999).
- Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2001).
- Rouse, J. & Jackson, S. P. Interfaces between the detection, signaling, and repair of DNA damage. *Science* **297**, 547–551 (2002).
- Callebaut, I. & Morrison, J. P. From BRCA1 to RAP1: A widespread BRCT module closely associated with DNA repair. *FEBS Lett.* **400**, 25–30 (1997).
- Sun, Z., Hsiao, J., Fay, D. S. & Stern, D. F. Rad53 FHA domain associated with phosphorylated Rad9 in the DNA damage checkpoint. *Science* **281**, 272–274 (1998).
- Vialard, J. E., Gilbert, C. S., Green, C. M. & Lowndes, N. F. The budding yeast checkpoint protein rad9 is subject to Mec1/Tel1-dependent hyperphosphorylation and interacts with Rad53 after DNA damage. *EMBO J.* **17**, 5679–5688 (1998).
- Emili, A. MEC1-dependent phosphorylation of Rad9p in response to DNA damage. *Mol. Cell* **2**, 183–189 (1998).
- Ahn, J. Y., Li, X., Davis, H. L. & Canman, C. E. Phosphorylation of threonine 68 promotes oligomerization and autophosphorylation of the CHK2 protein kinase via the forkhead-associated domain. *J. Biol. Chem.* **277**, 19389–19395 (2002).
- Xu, X., Tsvetkov, L. M. & Stern, D. F. Chk2 activation and phosphorylation-dependent oligomerization. *Mol. Cell Biol.* **22**, 4419–4432 (2002).
- Scully, R. *et al.* Dynamic changes of BRCA1 subnuclear localization and phosphorylation state are initiated by DNA damage. *Cell* **90**, 425–435 (1997).
- Ward, I. M., Wu, X. & Chen, J. Threonine 68 of Chk2 is phosphorylated at the site of DNA strand breaks. *J. Biol. Chem.* **276**, 47755–47758 (2001).
- Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. *J. Cell Biol.* **146**, 905–916 (1999).
- Paull, T. T. *et al.* A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10**, 886–895 (2000).
- Matsuoka, S., Huang, M. & Elledge, S. J. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**, 1893–1897 (1998).
- Melchionna, R., Chen, X. B., Blasina, A. & McGowan, C. H. Threonine 68 is required for radiation-induced phosphorylation and activation of Cds1. *Nature Cell Biol.* **2**, 762–765 (2000).
- Matsuoka, S. *et al.* Ataxia telangiectasia-mutated phosphorylates Chk2 *in vivo* and *in vitro*. *Proc. Natl Acad. Sci. USA* **97**, 10389–10394 (2000).

21. Ahn, J.-H., Schwarz, J. K., Piwnicka-Worms, H. & Canman, C. E. Threonine 68 phosphorylation by ataxia telangiectasia-mutated is required for efficient activation of Chk2 in response to ionizing radiation. *Cancer Res.* **60**, 5934–5936 (2000).
22. Chaturvedi, P. *et al.* Mammalian Chk2 is a downstream effector of the ATM-dependent DNA damage checkpoint pathway. *Oncogene* **18**, 4047–4054 (1999).
23. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
24. Painter, R. B. & Young, B. R. Radiosensitivity in ataxia-telangiectasia: A new explanation. *Proc. Natl Acad. Sci. USA* **77**, 7315–7317 (1980).
25. Falck, J., Mailand, N., Syljuasen, R. G., Bartek, J. & Lukas, J. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature* **410**, 842–847 (2001).
26. Hirao, A. *et al.* DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science* **287**, 1824–1827 (2000).
27. Hirao, A. *et al.* Chk2 is a tumour suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. *Mol. Cell. Biol.* **22**, 6521–6532 (2002).
28. Jack, M. T. *et al.* Chk2 is dispensable for p53-mediated G1 arrest but is required for a latent p53-mediated apoptotic response. *Proc. Natl Acad. Sci. USA* **99**, 9825–9829 (2002).
29. Chehab, N. H., Malikzay, A., Apple, M. & Halazonetis, T. D. Chk2/hCds1 functions as a DNA damage checkpoint in G1 by stabilizing p53. *Genes Dev.* **14**, 278–288 (2000).
30. Shieh, S. Y., Ahn, J., Tamai, K., Taya, Y. & Prives, C. The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev.* **14**, 289–300 (2000).

Acknowledgements We thank L. Wang for technical support. We also thank Mayo Protein Core facility for synthesis of peptides. We are grateful to L. Karnitz and S. Kaufmann and members of the Chen and Karnitz laboratories for discussions and ongoing technical support. This work is supported in part by grants from the National Institute of Health, the Prospect Creek Foundation and the Breast Cancer Research Foundation. J.C. is a recipient of a DOD breast cancer career development award.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C. (e-mail: chen.junjie@mayo.edu).

MDC1 is a mediator of the mammalian DNA damage checkpoint

Grant S. Stewart*, Bin Wang*, Colin R. Bignell†, A. Malcolm R. Taylor† & Stephen J. Elledge*‡§

* Verna & Mars McLean Department of Biochemistry and Molecular Biology, ‡ Department of Molecular and Human Genetics, and § Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA
† CRC Institute for Cancer Studies, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

To counteract the continuous exposure of cells to agents that damage DNA, cells have evolved complex regulatory networks called checkpoints to sense DNA damage and coordinate DNA replication, cell-cycle arrest and DNA repair¹. It has recently been shown that the histone H2A variant H2AX specifically controls the recruitment of DNA repair proteins to the sites of DNA damage^{2–4}. Here we identify a novel BRCA1 carboxy-terminal (BRCT) and forkhead-associated (FHA) domain-containing protein, MDC1 (mediator of DNA damage checkpoint protein 1), which works with H2AX to promote recruitment of repair proteins to the sites of DNA breaks and which, in addition, controls damage-induced cell-cycle arrest checkpoints. MDC1 forms foci that co-localize extensively with γ -H2AX foci within minutes after exposure to ionizing radiation. H2AX is required for MDC1 foci formation, and MDC1 forms complexes with phosphorylated H2AX. Furthermore, this interaction is phosphorylation dependent as peptides containing the phosphorylated site on H2AX bind MDC1 in a phosphorylation-dependent manner. We have shown by using small interfering RNA (siRNA) that cells lacking MDC1 are sensitive to ionizing radiation, and that MDC1 controls the formation of damage-induced 53BP1,

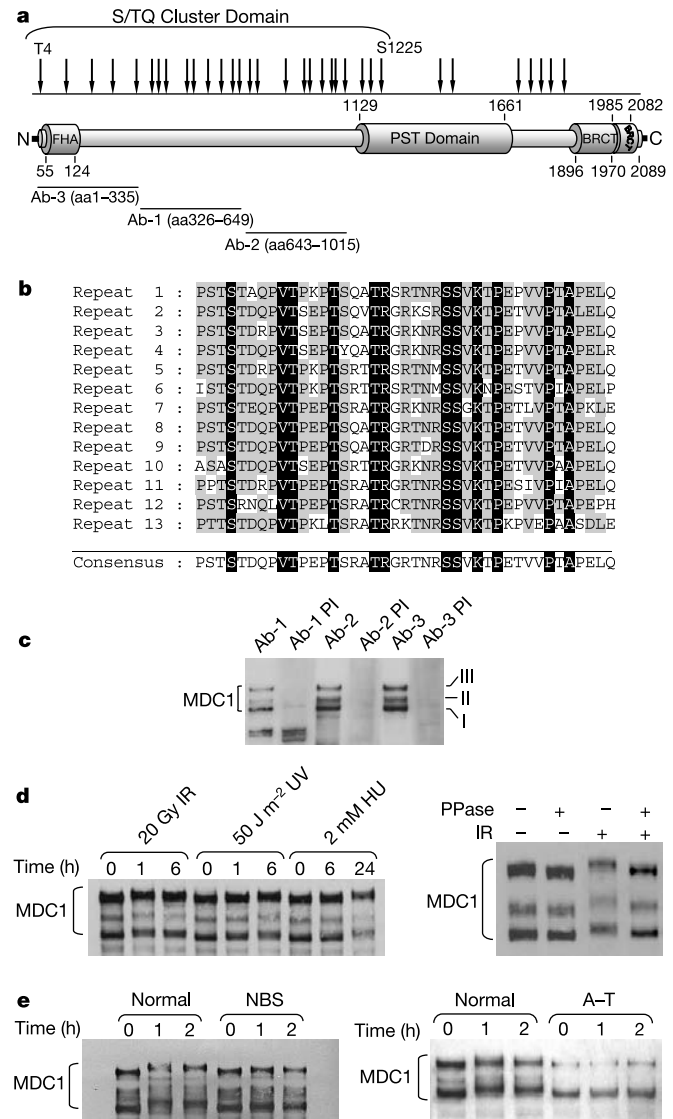


Figure 1 MDC1 is phosphorylated in response to DNA damage and DNA replication stress. **a**, A diagrammatic representation of the MDC1 protein. The amino acids encompassing each domain are indicated. PST indicates proline/serine/threonine repeat domain. The red arrows indicate potential phosphatidylinositol-3-OH kinase-like kinase phosphorylation motifs (SQ/TQ). The fragments of the MDC1 protein used to make anti-MDC1 antibodies (Ab-1, Ab-2, Ab-3) are indicated. **b**, Alignment of the 41-amino-acid repeat sequence that composes the PST domain. Black shaded boxes indicate conserved amino acids and grey boxes indicate similar amino acids. **c**, Recognition of three isoforms (I, II and III) of human MDC1. PI refers to pre-immune serum. **d**, DNA damage-induced phosphorylation of MDC1. Cells were treated with 20 Gy IR, 50 J m⁻² UV or 2 mM HU and harvested at the times indicated. IR-treated cell extracts were also incubated with and without λ protein phosphatase (λ PPase). **e**, MDC1 is phosphorylated in response to IR in an ATM- and Nbs1-dependent manner. Normal, NBS and A-T lymphoblasts were irradiated with 20 Gy of IR and harvested at the times indicated.

BRCA1 and MRN foci, in part by promoting efficient H2AX phosphorylation. In addition, cells lacking MDC1 also fail to activate the intra-S phase and G2/M phase cell-cycle checkpoints properly after exposure to ionizing radiation, which was associated with an inability to regulate Chk1 properly. These results highlight a crucial role for MDC1 in mediating transduction of the DNA damage signal.

Mediators are an emerging class of checkpoint proteins involved in transducing the DNA damage signal. The prototypical mediator

is the Rad9 protein in *Saccharomyces cerevisiae*. Rad9 is phosphorylated by the ATR (ataxia–telangiectasia and Rad3 related) homologue, Mec1, in response to DNA damage and controls the activation of the Chk1 and Chk2 (scRAD53) homologues^{5,6}. Rad9 contains two BRCT motifs that are required for its checkpoint functions⁷. No clear mammalian homologue of *S. cerevisiae* Rad9 has been identified. However, several large BRCT-repeat-containing proteins, such as TopBP1, 53BP1 and BRCA1, which have been implicated in the cellular response to DNA damage, may compose this class of mediator protein in higher eukaryotic cells^{8–11}.

To search for proteins involved in various aspects of DNA repair, we investigated proteins with motifs common to DNA damage response proteins and found a protein (Kazusa DNA Research Institute clone KIAA0170) that possesses the characteristics of a DNA damage mediator and named it MDC1. MDC1 contained two carboxy-terminal BRCT-repeats and an amino-terminal phospho-amino-acid-binding motif called a FHA domain, which is also found in several DNA damage response proteins such as Chk2, Rad53, Cds1 and Nbs1 (refs 12–15). A large S/TQ cluster domain can be found encompassing the N-terminal half of the protein (Fig. 1a). Both BRCA1 and Chk2 have S/TQ cluster domains and are phosphorylated within these domains by ATM (ataxia–telangiectasia mutated) and ATR after DNA damage^{16–19}. MDC1 also contains a

large central proline/serine/threonine-rich repeat (PST) domain that has no significant homology to any other protein in the database (Fig. 1b).

To investigate the role of MDC1 in the DNA damage response, three polyclonal antisera generated to non-overlapping fragments of the MDC1 open reading frame (ORF) all specifically recognized three bands (I, II and III) of approximately 250 kDa, which were not recognized by the pre-immune sera (Fig. 1c). All three bands diminished in intensity when cells were pre-treated with three different MDC1-specific siRNA oligonucleotides but not control siRNAs demonstrating specificity (data not shown and see Fig. 5a). To address whether the expression or mobility of MDC1 was affected by DNA damage, cells were exposed to a variety of genotoxic agents and blotted for MDC1. All isoforms demonstrated a reduced mobility after exposure of cells to ionizing radiation (IR), ultraviolet radiation (UV) and hydroxyurea (HU) (Fig. 1d). The altered mobility was ablated by phosphatase treatment, indicating that MDC1 is modified by phosphorylation after DNA damage (Fig. 1d). Furthermore, the phosphorylation of MDC1 after IR is dependent on the presence of functional ATM and Nbs1 (Fig. 1e). Both ATM and ATR could be shown to phosphorylate a fragment of the MDC1 protein encompassing the S/TQ cluster domain *in vitro* (data not shown).

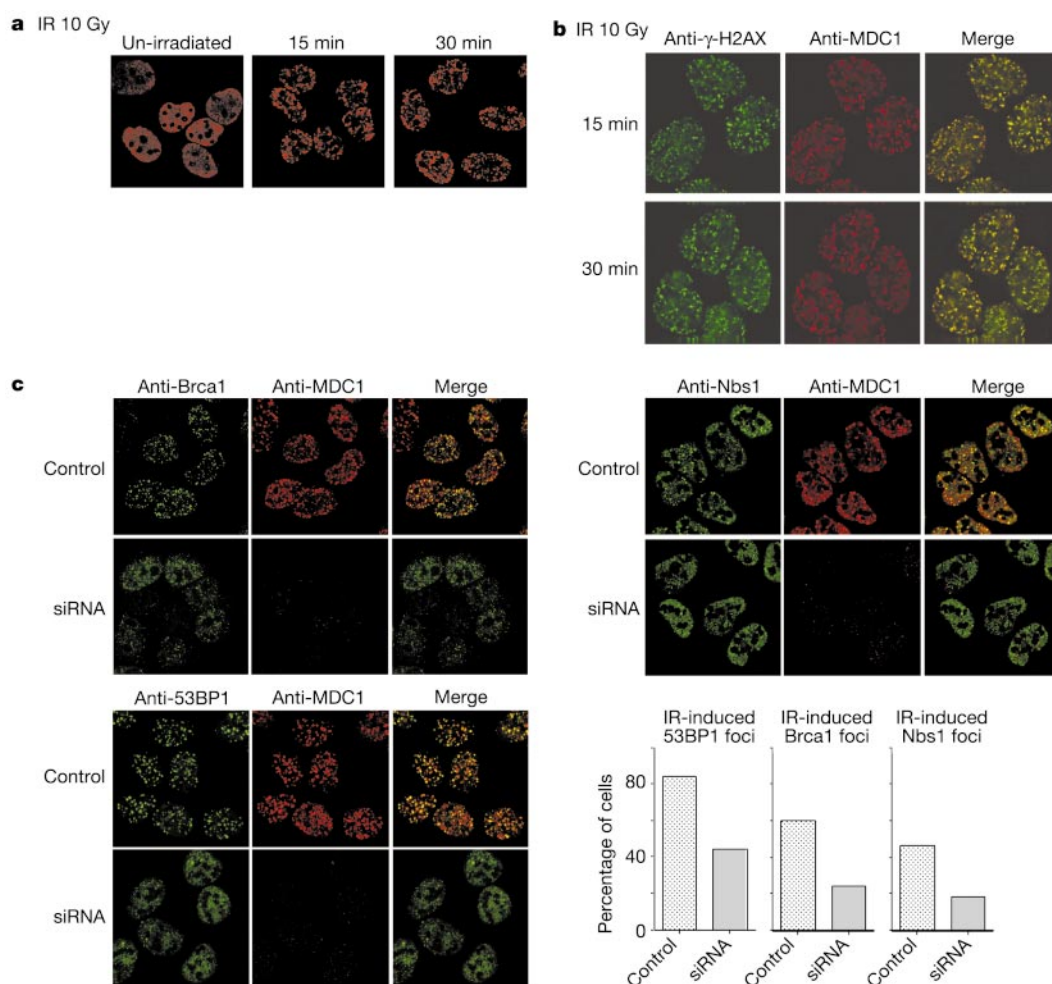


Figure 2 MDC1 regulates BRCA1, 53BP1 and Nbs1 foci formation. **a**, IR-induced MDC1 foci formation. Cells were untreated or irradiated with 10 Gy, fixed and stained with anti-MDC1 antibodies at the times indicated. **b**, Co-localization of MDC1 and γ -H2AX foci after 10 Gy IR. **c**, Inhibition of MDC1 results in defective BRCA1, 53BP1 and Nbs1 foci formation after IR exposure. U2OS cells were transfected with a control or MDC1 siRNA and irradiated with 10 Gy IR. Cells were fixed at 2 h post-irradiation and stained with the

indicated antibodies. Cells were fixed at 6 h post-irradiation to visualize Nbs1 foci. The percentages of cells with the respective foci are indicated. When cells were treated with MDC1 siRNA, the percentage of cells listed refers to the percentage of cells that lack detectable MDC1 but contain the indicated foci. Images were taken with a Zeiss confocal microscope.

Many DNA damage-signalling proteins are recruited to sites of damage²⁰. It has been suggested that the order and timing of these events are critical for normal DNA repair. To assess whether MDC1 could also localize to sites of damage, cells were treated with IR and stained with anti-MDC1 antiserum. MDC1 rapidly formed foci in over 95% of cells within 15 min after exposure to IR (Fig. 2a). A proportion of un-irradiated cells also contained MDC1 foci, indicating that MDC1 may be responding to endogenous damage or replication stress. No foci were observed in the cells stained with the pre-immune serum (Supplementary Information Fig. 1). MDC1 also formed foci after UV-irradiation (data not shown) indicating that MDC1 can respond to multiple types of DNA aberrations.

A time-dependent, sequential assembly of repair factors at the sites of damage has been demonstrated²⁰; γ -H2AX (the phosphorylated form of H2AX) foci appear within minutes after irradiation. 53BP1, BRCA1 and Mre11/NBS1/hRad50 complex are subsequently recruited to γ -H2AX-positive repair foci. To determine the kinetics of MDC1 foci compared with other foci-forming proteins, co-localization studies were done. Strikingly, within 15 min of exposure to ionizing radiation, MDC1 foci extensively overlap with γ -H2AX foci and remain co-localized throughout the time course (Fig. 2b). MDC1 foci also significantly co-localize with 53BP1 and BRCA1 foci at 2 h post-irradiation, and at later times also exhibited co-localization with Nbs1 foci (Fig. 2c). The numbers of MDC1 foci per cell were in excess of 53BP1, BRCA1 and Nbs1 foci at corresponding times.

Several proteins that show co-localization in foci exist as pre-assembled complexes²¹. Therefore, we determined whether MDC1 could bind to the components of the MRN complex and other known damage response proteins. MDC1 could be shown to co-precipitate with Nbs1, hRad50 and hMre11 before and after DNA damage and to a lesser extent with ATM and the FANCD2 protein (Fig. 3). The prior exposure of cells to IR appears to reduce the interaction of MDC1 with these proteins slightly. Given that MDC1

binds tightly to chromatin after DNA damage (data not shown), it is likely that the reduction of co-immunoprecipitated protein after exposure to IR is due to a reduced ability to efficiently solubilize MDC1 protein before immunoprecipitation. MDC1 could also be shown to interact strongly with 53BP1 and SMC1 in a manner unaffected by DNA damage. It is therefore likely that MDC1 exists within a large complex consisting of several DNA repair proteins and cell-cycle checkpoint regulatory proteins.

There appears to be a hierarchy of proteins involved in the assembly of damage-responsive foci. H2AX is required for the formation of 53BP1, NBS1 and BRCA1 foci^{10,22,23}. To determine whether MDC1 functions in this pathway, we asked if MDC1 was required for the ability of these repair proteins to form foci. Cells were treated with a control siRNA or MDC1 siRNA, irradiated and stained for 53BP1, BRCA1 and Nbs1 foci. Cells lacking MDC1 showed a significant reduction in the number of cells with Nbs1 foci, with the number of residual foci in each cell also being reduced (Fig. 2c). This suggests that MDC1 is required for efficient localization of the MRN complex after DNA damage. In addition, both 53BP1 and BRCA1 foci were similarly affected in the MDC1-siRNA-treated cells compared with the control siRNA-treated cells (Fig. 2c). These data demonstrate that MDC1 is likely to function upstream of 53BP1, BRCA1 and the MRN complex, possibly acting as a scaffold for protein recruitment. Consistent with this hypothesis, MDC1 foci were not affected in NBS or ATLD cells or when cells were treated with 53BP1-specific siRNA (data not shown). It should be noted that inhibition of MRN and 53BP1 foci was not complete, suggesting that either the siRNA treatment leaves residual functional MDC1 capable of limited foci formation, or that a partly redundant pathway exists to localize these proteins.

Because MDC1 and H2AX extensively co-localize and display similar kinetics of foci formation after IR, we examined the ability of MDC1 to form foci in the absence of H2AX and other checkpoint proteins. Mouse MDC1 showed a punctate nuclear staining pattern in wild-type mouse embryo fibroblasts (MEFs), which re-organized into large foci in response to DNA damage. MDC1 foci failed to form in H2AX null MEFs (Fig. 4a) but did form foci in Chk2 null MEFs (data not shown). Thus, H2AX acts upstream of MDC1. Surprisingly, the phosphorylation of MDC1 is also partly dependent on the presence of H2AX (Fig. 4b) but Chk2 loss has little or no effect on phosphorylation. Given that one of the functions of H2AX is to specifically recruit DNA repair proteins to sites of damage^{2,3}, this suggests that the modification of MDC1 by ATM and other kinases might be enhanced by recruitment to sites of DNA damage. In addition, the damage-induced phosphorylation of 53BP1 was also found to be partly dependent on H2AX (Fig. 4b), consistent with its interaction with MDC1. This contrasts with the ATM-dependent phosphorylation of Nbs1, which occurs normally in the absence of H2AX (ref. 2).

The fact that H2AX and MDC1 form completely overlapping foci with identical kinetics suggests that they might have a more intimate relationship. To examine this we immunoprecipitated MDC1 and tested for the presence of phospho-S139 H2AX. Phosphorylated H2AX co-immunoprecipitated with MDC1 and the association was increased in the presence of DNA damage (Fig. 4c). The presence of γ -H2AX in undamaged cells may represent intrinsic cellular damage. It is thought that the phosphorylation of H2AX on S139 triggers its ability to recruit factors such as MDC1 to foci. To examine this we used an S139-phosphorylated and unphosphorylated peptide from the C-terminal tail of H2AX coupled to beads to assess whether MDC1 could be pulled down from cellular extracts. MDC1 bound specifically to the phosphorylated H2AX peptide and not to the unphosphorylated H2AX peptide or a control-phosphorylated peptide from cyclin E (Fig. 4d). In addition to MDC1, 53BP1 was also pulled down by the phospho-H2AX peptide, confirming the previously published association of H2AX and 53BP1²³. Not all DNA repair/checkpoint proteins were found to

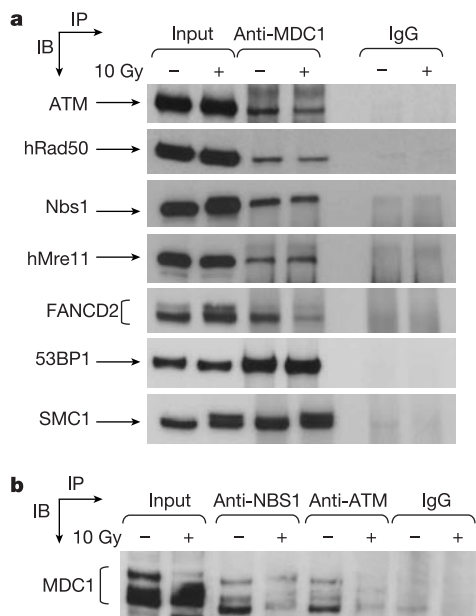


Figure 3 MDC1 associates with DNA damage checkpoint proteins. Cells were mock irradiated (—) or treated with 10 Gy IR (+) and harvested 1 h later. Immunoprecipitations (IP) or immunoblots (IB) were done with the antibody indicated or a non-specific, species-matched IgG control. Input represents 1% of total protein used in IPs. The unavailability of immunoprecipitating antibodies to 53BP1, SMC1 and FANCD2 prevented examination of the reciprocal co-immunoprecipitation with MDC1; however, the interaction was confirmed by using the two other anti-MDC1 antibodies generated to non-overlapping epitopes (data not shown).

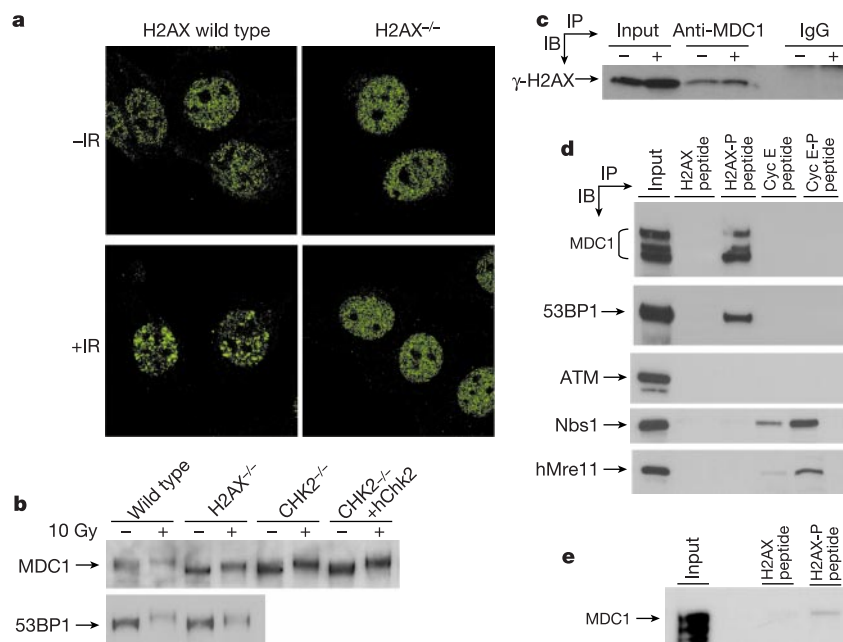


Figure 4 Interactions between MDC1 and H2AX. **a**, MDC1 foci formation is H2AX dependent. H2AX^{-/-} MEFs and the corresponding wild-type MEFs were exposed to IR, fixed and stained with anti-MDC1 antibodies at 30 min post-irradiation. **b**, MDC1 and 53BP1 phosphorylation is dependent on H2AX. Wild type, H2AX^{-/-} MEFs, CHK2^{-/-} and human CHK2 complemented CHK2^{-/-} MEFs were irradiated and harvested after 1 h. **c**, Association of MDC1 with γ -H2AX. Immunoprecipitations (IP) and immunoblots (IB) were done with the antibody indicated or a non-specific, species-matched IgG control.

d, A phosphorylated H2AX peptide recruits MDC1. Cell extracts were incubated with a peptide from the C-terminal tail of H2AX. (P) indicates phosphorylation of the peptide on serine-139. A cyclin E peptide and its phosphorylated version (P) were used as specificity controls. **e**, *In vitro* binding of MDC1 to the phosphorylated H2AX peptide. Phosphorylated or unphosphorylated H2AX-coupled peptide was added to ³⁵S-methionine-labelled, *in vitro* translated MDC1 and incubated for 1 h. Peptide-bound proteins were detected by using autoradiography.

associate with H2AX as components of the MRN complex: ATM and hChk2 (data not shown) were not significantly enriched by the H2AX p-S139 peptide. As we previously detected Nbs1 and ATM in MDC1 IPs, we had expected to recover these proteins with MDC1; however, it is possible that these proteins exist in a complex with MDC1 that already has γ -H2AX in it or that the complex containing MRN and ATM cannot bind in the context of the bead for steric reasons.

To further assess the interaction of MDC1 and the H2AX phospho-peptide, MDC1 was translated *in vitro* by using a rabbit reticulocyte lysate system and incubated with the H2AX-peptide-coupled beads. The MDC1 translated *in vitro* was pulled down specifically by the H2AX phospho-peptide, again demonstrating that the interaction is phospho dependent (Fig. 4e). However, it is unclear whether this interaction is direct as the reticulocyte lysates may contain bridging proteins.

To examine the genetic role of MDC1 in the DNA damage response, siRNA was used to deplete MDC1. Cells were transiently transfected with either a control siRNA or one of three MDC1-specific siRNAs, harvested 48–72 h after transfection, and their protein expression determined. All three MDC1-specific siRNA oligonucleotides decreased by more than 80% in the overall MDC1 protein expression, compared with the mock or control siRNA-transfected cells (Fig. 5a). To examine the ability of MDC1-depleted cells to respond to damage, siRNA-transfected cells were plated at low density, irradiated with IR and assessed for their ability to form colonies. Cells with reduced MDC1 exhibited hypersensitivity to the killing effects of IR (Fig. 5b) when compared with the control cells. Furthermore, transfection with MDC1-specific siRNA also reduced the number of colonies formed from undamaged cells when compared with control cells, indicating that MDC1 function may be required to maintain cell viability.

The sensitivity to IR suggests an important role in responding to DNA damage; therefore we checked the integrity of cell-cycle

checkpoints in these cells. To test whether MDC1 could function in regulating the intra-S-phase checkpoint, cells were either transfected with control siRNA or MDC1-specific siRNA oligonucleotides. After 48 h, cells were exposed to 15 Gy of IR and the rate of incorporation of tritiated thymidine was determined. Control cells exhibited a decrease of approximately 50% in the rate of DNA synthesis. In contrast, cells treated with MDC1 siRNA oligonucleotides 4 or 5 only decreased the rate of DNA synthesis by 25% (Fig. 5c). This demonstrates a role for MDC1 in regulating S-phase progression after DNA damage.

We next examined the integrity of the G2/M checkpoint. Cells treated with MDC1-specific siRNA were irradiated and labelled with an anti-phospho-histone H3 antibody as a marker of mitotic cells. A clear reduction in phospho-H3-positive cells was observed in the control-treated cells after IR exposure, whereas a significant number of the cells lacking MDC1 entered mitosis (Fig. 5d), indicative of a defect in the ability to arrest the cell cycle in G2 phase.

To gain insight into the precise mechanism by which MDC1 mediates both the S-phase and G2/M damage-activated cell-cycle checkpoints, cells were treated with MDC1-targeted siRNA and assessed for their ability to properly phosphorylate key effector molecules known to be critical for efficient checkpoint activation after DNA damage. We observed decreased phosphorylation of SMC1 S996 (more pronounced at 2 h) and Chk1 S345 in response to IR, and SMC1, Chk1 and RPA2 in response to UV (Fig. 5e). Nbs1 and Chk2 were not affected although we observed a small decrease in Chk2 T68 phosphorylation in response to UV in some experiments (data not shown). Chk1 is required for the G2/M checkpoint arrest^{24,25} and this defect is likely to explain the checkpoint defect we observed. Both SMC1^{26,27} and Chk1²⁸ have been implicated in the intra-S-phase checkpoint, and their defective regulation is likely to explain these defects as well. The fact that UV-induced signalling is impaired suggests that the ATR pathway depends in part upon MDC1 function, as ATR is the major regulator of Chk1 phosphoryl-

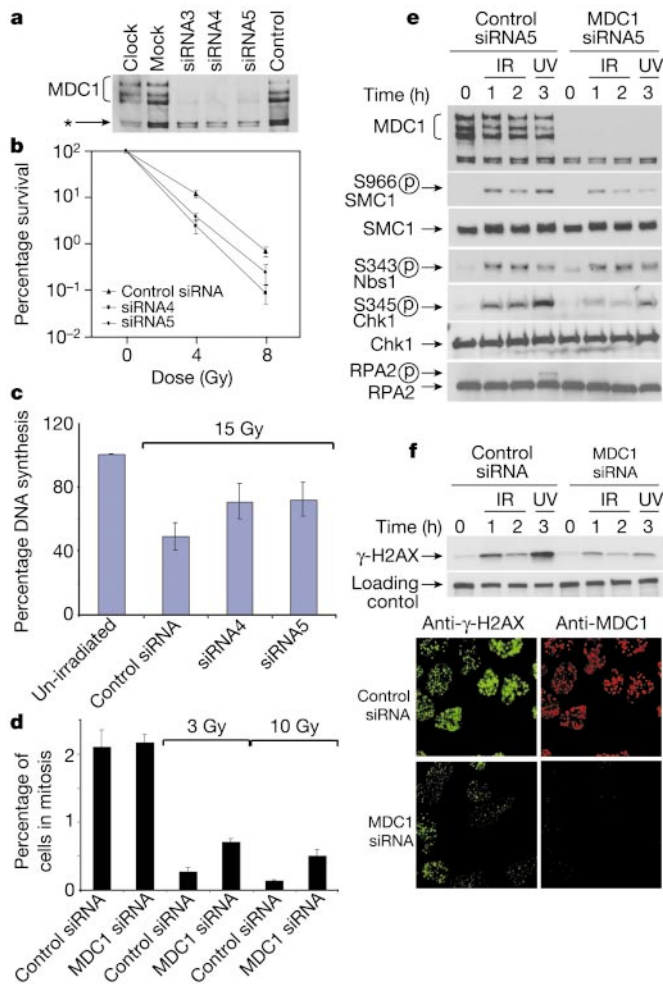


Figure 5 MDC1 inhibition results in defective IR-induced checkpoints and Chk1 and H2AX phosphorylation. **a**, siRNA-mediated reduction of MDC1 protein expression. Cells were transfected with a control or MDC1 siRNA and harvested at 48 h. Three different siRNAs to MDC1 were used. An asterisk indicates a cross-reacting band used as a loading control. **b**, Reduced MDC1 expression results in IR sensitivity. Control or MDC1 siRNA-treated cells were plated at low density, irradiated and colonies counted after 14 days. **c**, MDC1 prevents RDS. DNA synthesis was assessed 30 min after 15 Gy IR in U2OS cells transfected with a control or MDC1 siRNA oligonucleotides. **d**, Analysis of the G2/M checkpoint. Cells were untreated or γ -irradiated as indicated, then incubated for 1 h at 37 °C before fixation. Mitotic cells were determined by phospho-histone H3 staining and flow cytometry. **e**, MDC1 is required for DNA damage checkpoint signalling. Cells were transfected twice with siRNAs and irradiated 72 h later. The immunoblot with SMC1 as a control for protein loading shows a reduced mobility in extracts from damaged cells rather than increased protein loaded on the gel. **f**, MDC1 is required for H2AX phosphorylation. Control or MDC1 siRNA-treated cells were irradiated and the levels of H2AX phosphorylation detected by western blot and immunofluorescence 1 h post-irradiation.

ation. Together, these results indicate that MDC1 is a central transducer of the DNA damage checkpoint signal.

As MDC1 controls the phosphorylation of several checkpoint-responsive proteins, we sought to examine whether it might have a role in H2AX phosphorylation. Depleting cells of MDC1 protein significantly affected the phosphorylation of H2AX after both IR and UV radiation and the formation of phospho-H2AX foci (Fig. 5f). Thus MDC1 and H2AX share a mutual dependency for phosphorylation and foci formation.

In this study, we have identified a new DNA damage checkpoint protein, MDC1, which transduces the DNA damage signal and shares an intimate relationship with H2AX. In response to DNA damage, MDC1 and H2AX exist in a complex, are phosphorylated

and co-localize to sites of DNA damage, all in a mutually dependent fashion. Both proteins are required for optimal formation of MRN, 53BP1 and BRCA1 foci. Furthermore, peptides representing the tail of H2AX specifically recruit MDC1 and 53BP1 protein in a phosphorylation-dependent manner, suggesting a model in which H2AX is phosphorylated in response to damage and the phospho-H2AX recruits MDC1 leading to its phosphorylation. However, we cannot determine whether MDC1 controls the initial phosphorylation of H2AX, or merely protects H2AX from dephosphorylation through formation of a complex with H2AX. It is possible that there is a self-reinforcing loop whereby MDC1 and H2AX mutually stimulate each other's phosphorylation and help larger complexes of proteins associate at these foci. Furthermore, this effect on H2AX phosphorylation may, in part, help to explain why the recruitment of 53BP1, BRCA1 and Nbs1 to damage foci partly depends on MDC1. The ability of MDC1 to form complexes with MRN and 53BP1 may also contribute to this dependency. Regardless, it is clear from the data presented here that H2AX and MDC1 are intimately related and work together to mediate the DNA damage response.

Although MDC1 and H2AX have many common features, there are some important differences. MDC1 is required for the intra-S-phase checkpoint and the G2/M DNA damage checkpoint. H2AX has recently been shown to have a mild G2/M defect in response to low doses of IR³, much less severe than that observed in MDC1. MDC1 exists in complexes with several checkpoint-signalling proteins such as the MRN complex, SMC1, 53BP1 and FANCD2. Furthermore, MDC1 is required for the efficient phosphorylation of several checkpoint-signalling proteins including SMC1, RPA2 and the crucial checkpoint kinase Chk1. The defect in Chk1 phosphorylation explains the defects in both the intra-S-phase and G2/M checkpoints. The defect in MDC1 checkpoint signalling is more pronounced in response to UV than IR. This, together with the regulation of Chk1, suggests that MDC1 plays an important role in the ATR signalling pathway that controls Chk1 phosphorylation²⁵ and a less important role in the ATM pathway. This may explain the mild defect in the IR-induced phosphorylation of SMC1, which is primarily ATM dependent. The mild SMC1 phosphorylation defect becomes more pronounced at later times in the absence of MDC1, consistent with the late role of ATR in IR-induced phosphorylation of substrates shared by ATM. MDC1 may play a role in the recruitment of checkpoint kinases such as ATR to substrates such as H2AX and the pathway controlling Chk1, which may include BRCA1 and Claspin. Therefore, MDC1 plays a critical role in the DNA damage checkpoint response. MDC1 represents the third member of the mediator family of proteins in mammals. Each of these proteins plays important and potentially overlapping roles in transducing the DNA damage signal to promote genomic stability. As a new member of this family, MDC1 is also likely to be involved in tumorigenesis. □

Methods

Cells

Lymphoblastoid cell lines (LCLs) were routinely maintained in RPMI medium supplemented with 10% foetal bovine serum (FBS), glutamine and penicillin and streptomycin (pen/strep). U2OS cells were maintained in McCoy's 5A medium supplemented with 10% FBS, glutamine and pen/strep. All MEF cell lines were maintained in DMEM medium supplemented with 10% FBS, glutamine and pen/strep.

siRNA

The siRNA duplexes were 21 base pairs with a 2-base deoxynucleotide overhang (Dharmacon Research). The sequences of MDC1 siRNA4 and siRNA5 oligonucleotides were GUCUCCAGAAAGACAGUGAdTdT and ACAGUUGUCCCCACAGCCdTdT, respectively. The control siRNA used was CGUACGCGGAUACUUCGAdTdT against LacZ. Cells were transfected with siRNA duplexes by using Oligofectamine (Invitrogen), following the manufacturer's instructions.

MDC1 antibody generation

A full-length MDC1 clone (KIAA0170) was generously provided by the Kazusa DNA Research Institute. Fragments of the cDNA were subcloned into a glutathione S-transferase (GST) expression vector (Amersham Biosciences), expressed in bacteria and

purified by using glutathione-coupled sepharose beads (Amersham Biosciences). Purified GST fusion proteins were injected into rabbits and the antisera were affinity purified by using the respective antigen (Bethyl Laboratories). Ab-1 was generated to a fragment of MDC1 encompassing amino acids 326–649, Ab-2 was generated to a fragment of MDC1 encompassing amino acids 643–1015 and Ab-3 was generated to a fragment of MDC1 encompassing amino acids 1–335.

Immunoblot analysis

Cells were sonicated in UTB buffer (8 M urea, 150 mM β -mercaptoethanol, 50 mM Tris/HCl pH 7.5) and cellular debris removed by centrifugation. Protein concentration was determined by using the BioRad Bradford Protein determination reagent. Proteins were fractionated in 6% SDS–polyacrylamide gels. Proteins were transferred to nitrocellulose, and immunoblots were performed by using the appropriate antibody. Antibodies to Nbs1, hMre11 and hRad50 were obtained from Genetex, the phospho-Nbs1 (serine-343) antibody was obtained from Cell Signaling and the anti-phospho-H2AX (serine-139) antibodies were obtained from Upstate Biotechnology. The anti-FANCD2 and anti-Chk1 antibodies were purchased from Santa Cruz. J. Qin provided antibodies to SMC1 and phosphorylated SMC1 (ref. 21). P. Carpenter supplied antibodies to 53BP1.

Immunoprecipitation and phosphatase treatment

LCLs ($\sim 4 \times 10^7$) were either mock irradiated or irradiated with 10 Gy of ionizing radiation and incubated for 1 h at 37 °C. The cells were then lysed for 30 min in NETN lysis buffer (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% NP40 supplemented with protease inhibitors (Roche) and Benzonase (Novagen)). The clarified extract was pre-cleared with the appropriate IgG (Dako) and protein A or G beads (Amersham Biosciences) for 1 h at 4 °C. Immunoprecipitating antibody (5 μ g) was added with protein A or G beads to the pre-cleared supernatant and incubated for 3 h at 4 °C. The immunocomplexes were washed four times in NETN lysis buffer (containing 0.5% NP40), boiled in SDS sample buffer, loaded on an SDS–polyacrylamide gel. Proteins were analysed by immunoblotting using standard methods and detected as described above. For phosphatase treatment of cell extract, 1,200 units of λ protein phosphatase (New England Biolabs) were added to the extract in the presence of MnCl₂ and incubated for 30 min at 30 °C. The phosphatase was then heat inactivated.

H2AX peptide pull-down assay

Two peptides to the last C-terminal 10 amino acids of human H2AX (CKATQASQEY) were synthesized (Bethyl), one of which was phosphorylated on the serine residue. The peptides were coupled to beads by using the Sulfolink kit (Pierce). 20 μ g of coupled peptide were used per pull-down. The assay was essentially identical to the immunoprecipitation protocol as described above.

H2AX peptide *in vitro* binding assay

MDC1 construct (2 μ g) was translated *in vitro* for 90 min at 30 °C by using the TNT-coupled reticulocyte lysate system (Promega) containing 20 μ Ci of ³⁵S-labelled methionine (Amersham Biosciences). Half of each reaction was added to a volume of NETN lysis buffer containing 1% NP40. Coupled H2AX peptide (20 μ g) was added and incubated for 1 h at 4 °C. The beads were washed thoroughly with NETN and the bound proteins analysed by western blotting.

Colony-forming assay

U2OS cells transfected with siRNA were seeded at low density and irradiated with various doses of ionizing or UV radiation. Cells were left for 14 days at 37 °C to allow colonies to form. Colonies were stained with 2% methylene blue/50% ethanol and counted. Colonies were defined as containing 50 or more cells.

Radio-resistant DNA synthesis assay

The RDS assay was done as described as before. Briefly, cells were transfected with siRNA oligonucleotides, and 4 h later were placed into McCoy's 5A medium containing 10 nCi of ¹⁴C-labelled thymidine (NEN Life Science Products Inc.) per millilitre overnight. The medium containing ¹⁴C-labelled thymidine was then replaced with normal McCoy's 5A medium, and the cells were incubated for another 24 h. Cells were irradiated, incubated for 30 min at 37 °C, and then pulse-labelled with 2.5 μ Ci [3H]thymidine (NEN Life Science Products Inc.) per millilitre for 15 min. Cells were harvested, washed twice with PBS, and fixed in 70% methanol for 30 min. After the cells were transferred to Whatman filters and fixed sequentially with 70% and then 90% methanol, the filters were air-dried and the amount of radioactivity was assayed in a liquid scintillation counter. The resulting ratios of ³H counts per minute to ¹⁴C counts per minute, corrected for those counts per minute that were the result of channel crossover, were a measure of DNA synthesis.

Immunofluorescence

Cells were fixed with 3% paraformaldehyde/2% sucrose for 10 min, permeabilized with 0.5% Triton X-100 solution, and then immunostained with primary antibodies against various proteins and the appropriate Alexa488- (green, Molecular Probes) and Cy3- (red) conjugated secondary antibodies (Amersham Biosciences). Images were taken with a Zeiss confocal microscope.

Received 12 December 2002; accepted 24 January 2003; doi:10.1038/nature01446.

1. Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
2. Bassing, C. H. *et al.* Increased ionizing radiation sensitivity and genomic instability in the absence of histone H2AX. *Proc. Natl Acad. Sci. USA* **99**, 8173–8178 (2002).
3. Celeste, A. *et al.* Genomic instability in mice lacking histone H2AX. *Science* **296**, 922–927 (2002).
4. Fernandez-Capetillo, O. *et al.* DNA damage-induced G(2)-M checkpoint activation by histone H2AX and 53BP1. *Nature Cell Biol.* **4**, 993–997 (2002).
5. Vialard, J. E., Gilbert, C. S., Green, C. M. & Lowndes, N. F. The budding yeast Rad9 checkpoint protein is subjected to Mec1/Tel1-dependent hyperphosphorylation and interacts with Rad53 after DNA damage. *EMBO J.* **17**, 5679–5688 (1998).
6. Gilbert, C. S., Green, C. M. & Lowndes, N. F. Budding yeast Rad9 is an ATP-dependent Rad53 activating machine. *Mol. Cell.* **8**, 129–136 (2001).
7. Soulier, J. & Lowndes, N. F. The BRCT domain of the *S. cerevisiae* checkpoint protein Rad9 mediates a Rad9-Rad9 interaction after DNA damage. *Curr. Biol.* **9**, 551–554 (1999).
8. Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *J. Cell Biol.* **151**, 1381–1390 (2000).
9. Venkitaraman, A. R. Functions of BRCA1 and BRCA2 in the biological response to DNA damage. *J. Cell Sci.* **114**, 3591–3598 (2001).
10. Wang, B., Matsuoka, S., Carpenter, P. & Elledge, S. J. 53BP1, a mediator of the DNA damage checkpoint. *Science* **298**, 1435–1438 (2002).
11. Yamane, K., Wu, X. & Chen, J. A DNA damage-regulated BRCT-containing protein, TopBP1, is required for cell survival. *Mol. Cell Biol.* **22**, 555–566 (2002).
12. Callebaut, I. & Mornon, J. P. From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Lett.* **400**, 25–30 (1997).
13. Ozaki, T. *et al.* NFB1/KIAA0170 is a novel nuclear transcriptional transactivator with BRCT domain. *DNA Cell Biol.* **19**, 475–485 (2000).
14. Huang, M. & Elledge, S. J. in *Cold Spring Harbor Symposia on Quantitative Biology* Vol. LXV, 1–9 (Cold Spring Harbor Laboratory Press, New York, 2000).
15. Durocher, D. & Jackson, S. P. The FHA domain. *FEBS Lett.* **513**, 58–66 (2002).
16. Matsuoka, S., Huang, M. & Elledge, S. J. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**, 1893–1897 (1998).
17. Cortez, D., Wang, Y., Qin, J. & Elledge, S. J. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science* **286**, 1162–1166 (1999).
18. Matsuoka, S. *et al.* Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proc. Natl Acad. Sci. USA* **97**, 10389–10394 (2000).
19. Tibbetts, R. S. *et al.* Functional interactions between BRCA1 and the checkpoint kinase ATR during genotoxic stress. *Genes Dev.* **14**, 2989–3002 (2000).
20. Paull, T. T. *et al.* A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10**, 886–895 (2000).
21. Wang, Y. *et al.* BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* **14**, 927–939 (2000).
22. Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J. Cell Biol.* **146**, 905–916 (1999).
23. Rappold, L., Iwabuchi, K., Date, T. & Chen, J. Tumor suppressor p53 binding protein 1 (53BP1) is involved in DNA damage-signaling pathways. *J. Cell Biol.* **153**, 613–620 (2001).
24. Takai, H. *et al.* Aberrant cell cycle checkpoint function and early embryonic death in Chk1^{-/-} mice. *Genes Dev.* **14**, 1439–1447 (2000).
25. Liu, Q. *et al.* Chk1 is an essential kinase that is regulated by Atr and required for the G₂/M DNA damage checkpoint. *Genes Dev.* **14**, 1448–1459 (2001).
26. Kim, S. T., Xu, B. & Kastan, M. B. Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes Dev.* **16**, 560–570 (2002).
27. Yazdi, P. T. *et al.* SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev.* **16**, 571–582 (2002).
28. Zhao, H., Watkins, J. L. & Piwnicka-Worms, H. Disruption of the checkpoint kinase 1/cell division cycle 25A pathway abrogates ionizing radiation-induced S and G2 checkpoints. *Proc. Natl Acad. Sci. USA* **99**, 14795–14800 (2002).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank A. Nussenzweig for the gift of the H2AX null MEFs and the corresponding wild-type MEFs. We thank P. Cooper, J. Chen and S. Jackson for sharing unpublished information and helpful discussions. We thank D. Lou for excellent technical assistance and N. Foray for technical advice. This work was supported by a National Institutes of Health Grant to S.J.E. B.W. is a fellow of the US Army Breast Cancer Postdoctoral Trainee Award, and S.J.E. is supported by a grant from the NIH, is an Investigator with the Howard Hughes Medical Institute and is a Senior Scholar of the Ellison Foundation. C.R.B. is supported by the Leukemia Research Fund. G.S.S. is supported by the European Molecular Biology Organization (EMBO).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.J.E. (e-mail: selledge@bcm.tmc.edu).

Talk to your analyst

Analytical instruments and offshoots, including a new way with images.

Automated weighing stations

Tecan www.tecan.com

Time-saving measures

Tecan's automated weighing systems can do many of the tasks associated with gravimetric analysis of samples. They can also be configured to perform dilution of solid samples with a wide variety of solvents. Samples may be weighed in either tubes or microtitre plates and subsequently diluted in a solvent of the user's choice. Applications include downstream assay or sample analysis in compound logistics and management, inventory control, chemical synthesis, high-throughput screening, combinatorial chemistry, ADMET (absorption, distribution, metabolism, excretion and toxicity) and other quality control functions.

Amicon Ultra centrifugal filters

Millipore www.millipore.com

For rapid recovery

These ultrafiltration centrifugal filters are designed for high-speed, high-recovery purification and concentration of proteins and nucleic acids. They combine a low-binding ultrafiltration membrane with a vertical design to deliver sample recoveries that are typically >90%. Sample volumes of <4–15 ml can be concentrated or diafiltered (desalt or buffer exchange) in as little as 15 minutes. The filters are compatible with all rotor types and allow direct pipettor sample access, eliminating the need for an extra processing step. Translucent housing with clearly marked filter devices and centrifugal filtrate tubes facilitates sample monitoring. The devices are available in packs of 8, 24 and 96 and are suitable for sample concentration or purification of dilute or purified protein solutes as well as complex samples such as fermentation broths or cell lysates.



Filter tip from Millipore: use Amicon Ultra.



Flexible spin: the Avanti takes centrifuge tubes of all shapes and sizes.

Avanti JE all-purpose centrifuge

Beckman Coulter www.beckmancoulter.com

Spin the bottle

The Avanti J-E compact general-purpose centrifuge system comes with 14 rotors to cover a wide range of bioseparation applications, including cellomics, genomics and proteomics. The centrifuge can be used with a variety of labware, including microtitre plates, 4 × 1 litre bottles and HarvestLine system liners. For clinical research, there are swinging bucket rotors that spin conical tubes and bottles of 15, 50 and 250 ml for blood component isolations with a range of gradient kits. A high-torque switch reluctance (SR) drive ensures rapid acceleration and deceleration and low heat output.

Pro-Pure

Amresco www.amresco-inc.com

Peroxide-free detergents

Tween 20, Triton X-100, Triton X-114 and Brij 35 make up Amresco's Pro-Pure line of peroxide-free detergents for protein isolation and sample preparation. They have less than 1 mg peroxide per ml and are packed under inert gas for stability. The detergents come as a 10-ml five-pack.

Optima 4300V

PerkinElmer www.perkinelmer.com/instruments

This one will test your metal

The Optima 4300V ICP-OES (inductively coupled plasma-optical emission spectroscopy) system can determine up to 72 elements in less than one minute. It is a radial-view instrument with a vertical-torch configuration and a solid-state RF generator that eliminates the need for consumable parts.

The system is suitable for analyses of difficult matrices such as oils/organics, precious metals and geochemical samples.

MicroAngelo

Asylum Research www.asylumresearch.com

Artistic licence

MicroAngelo is a standard feature of Asylum's MFP-3D atomic force microscope for nanolithography and nanomanipulation. It consists of a nanopositioning system, digital controller and Igor Pro software. To make the image below, a JPEG of Picasso's *Don Quixote* was imported into MicroAngelo. Shown is the AFM phase image of the nanolithographically patterned polycarbonate, 5- μ m scan. Users can import curves from a variety of programs or create their own mathematically designed



Can you see what it is yet? MicroAngelo at work.

curves and arrays. All data types are available for plotting during lithography, including deflection, lateral force, amplitude and phase. Imported images are converted to contour lines that are then scaled and used to drive the scanner to create the lithographic image. The scanner can be manipulated using a variety of pointing devices, including a mouse or a stylus pad. Applications include surface scratching and patterning, nanotube and particle manipulation and single-molecule experiments.

Flow Check

Polysciences www.polysciences.com
Bits and pieces for cytometer calibration

Polysciences has expanded its line of flow cytometry particles and sets for flow cytometer calibration. Flow Check flow cytometry particles are 6 µm in size, supplied at a concentration of 1.0×10^7 particles per ml in aqueous suspension, and fluorescently labelled with either FITC, phycoerythrin (PE), sulforhodamine, rhodamine or allophycocyanin (APC). Flow Check intensity particles are available in four levels of fluorescent intensity: lo-brite, med-brite, high-brite and maxi-brite. A colour compensation set containing maxi-brite particles labelled with FITC, PE, sulforhodamine and APC is also available. All intensity and colour compensation sets contain a bottle of undyed 6-µm particles at 1.0×10^7 particles per ml for comparison. Calibration particles are supplied as 5 ml aqueous suspensions in 10 ml opaque dropper bottles.

LyoPro 6000

Jouan Nordic www.heto-holten.com
Compact freeze dryers

Jouan Nordic now offers two models of freeze dryer. The LyoPro 6000-55 for general laboratory applications has a standard con-

denser temperature of -55°C while the more sophisticated LyoPro 6000-90, for applications using solvents other than water, employs a cascade refrigeration system to provide temperatures down to -90°C . Although compact, the freeze dryers provide an ice trapping capacity of 6 kg/24 hours, or 10 kg in total. A microprocessor is included for programming temperature, vacuum and alarm parameters.

Low Cross Talk Plate

Nunc www.nuncbrand.com
Enhance scintillation proximity assays

This 384-well polypropylene plate is lightweight, highly reflective and has low binding. It is especially suitable for luminescence applications, in particular scintillation proximity assays, but can also be used for fluorescence assays. Polypropylene plates have inherent hydrophobic properties and rounded square wells to minimize capillary action (wicking). The Low Cross Talk plate is compatible with automated equipment such as stackers, work stations and liquid dispensers.

LAS-3000 image analyser

Fuji Photo Film <http://home.fujifilm.com>
Camera happy

The latest incarnation of Fujifilm's ultra-sensitive Luminescent Image Analyzer series, the LAS-3000, features electronically cooled CCD camera technology and a simplified user interface. The LAS-3000 produces ultra-high-resolution image files of up to 6.3M pixels. Combined with Fujinon's highly sensitive lens and specially designed binning process, this allows the capture of chemiluminescent, fluorescent and gel documentation images of samples up to 25×25 cm in size with high sensitivity without sacrificing image resolution. The long exposure time and 16-bit data handling allow for sensitive and quantitative image analysis. Four kinds of light sources — blue epi-illuminator, white epi-illuminator, UV transilluminator and white transilluminator — and up to five optical filters including custom-made ones widen the sphere of imaging applications.

CMOSens SHT71

Sensirion www.sensirion.com
Plug-in digital humidity sensor

The SHT71 combines two sensor elements, an A/D converter, calibration and a digital interface on one chip. The module is supplied in a pin type package for 1.27 mm sockets with a thin bridge between the pins and the sensor to maximize thermal isolation and minimize interference from other circuits. Response time is less than 4 seconds. Each module is individually calibrated in a 'thunder scientific' precision humidity chamber and supplies a digital signal for



Sensirion SHT71: it's not big.

humidity and temperature, from which the dew point can also be calculated. This feature, in combination with the pin-type packaging, provides interchangeability of the sensor without requiring the user to recalibrate the entire system after changing the sensor. Applications include test and measurement, transmitters, data logging and climate control/building control HVACR.

These notes are compiled in the Nature office from information provided by the manufacturers. Press releases to be considered for publication can be sent to newproducts@nature.com.



LyoPro: freeze dryers on wheels.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakakamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Ground control

For many PhDs, the job outlook is increasingly grim. Unemployment among chemists in the United States has almost doubled over the past year, and biomedical research is looking less stable. But there is one discipline that is holding its own — Earth and space sciences. Unemployment among new PhDs in this field has fallen over the past five years, according to a recently released study by the American Geophysical Union, the American Geological Institute and the American Institute of Physics.

Despite the fact that many of the survey's responses were collected during the general economic downturn that followed the terrorist attacks in September 2001, the outlook for geoscientists still seems to be improving. In 1996, 59% of the survey's respondents considered their job outlook to be "hopeless" or "bad". In 2001, that figure was just 21%.

But some cracks are inevitably starting to show. Although salaries for most posts have seen steady growth, starting salaries for geoscience postdocs have not changed for two years. And the number of people doing postdocs, rather than securing permanent positions, is rising. Hints, perhaps, that the market may be slowing down or that federal grants are becoming harder to acquire.

Some geoscientist PhDs manage to secure better pay for one simple reason — experience. About a fifth of new PhDs in the field are awarded to people who are over 40 — a group that tends to be paid more and is less likely to pursue a postdoc. On average, geoscientists wait the longest before returning to university after they graduate, according to the US National Science Foundation. During these 'gap' years, they tend to gain valuable field experience, as well as clarifying their area of specialization. The fact that this approach seems to pay off begs the question of why graduates in other disciplines don't try it.

Paul Smaglik
Naturejobs editor



Contents

CAREERS AND RECRUITMENT

The merging world
of biophysics p970

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

The data busters

Making sense of the reams of information streaming out of genome projects requires a sophisticated blend of biology and physics, says Kendall Powell.

Where worlds collide: the University of California, San Diego, is making a concerted effort to bridge the gaps between biology and physics.

Fresh approaches to complex biological systems and the need to organize and make use of the mountains of data coming out of genomics have created an unprecedented demand for scientists who can straddle the interface between physics and the life sciences.

Physicists and biologists will have to work alongside each other to clear this century's highest scientific hurdles, even though cultural differences and old-fashioned attitudes can cause friction. To accelerate collaboration and train the new interdisciplinarians, many universities are bringing theoreticians and experimentalists together. Large numbers of interdisciplinary departments and graduate programmes are now starting up at universities around the world.

At the Center for Theoretical Biological Physics at the University of California, San Diego, co-director José Onuchic is delighted with the success of the centre's precursor programme, La Jolla Interfaces in

Science, which trains postdocs and graduate students to apply quantitative methods to biology. "We were told: 'the students won't be real physicists, chemists or biologists and they're not going to get any jobs,'" he says. "But it was exactly the opposite."

Nearly two-thirds of the postdocs on the programme have gone on to an assistant professorship, says Onuchic. And many of the rest have moved into high-paying positions in industry working in drug discovery or materials science.

With US\$10.5 million in funding spread over five years from the National Science Foundation, the centre fosters the development of physical approaches to biological systems and new forms of bio-inspired physics. It was set up with the goal of creating a two-way exchange between theoretical and molecular biology groups. To achieve this, each graduate student or postdoc is attached to two laboratories in different disciplines.

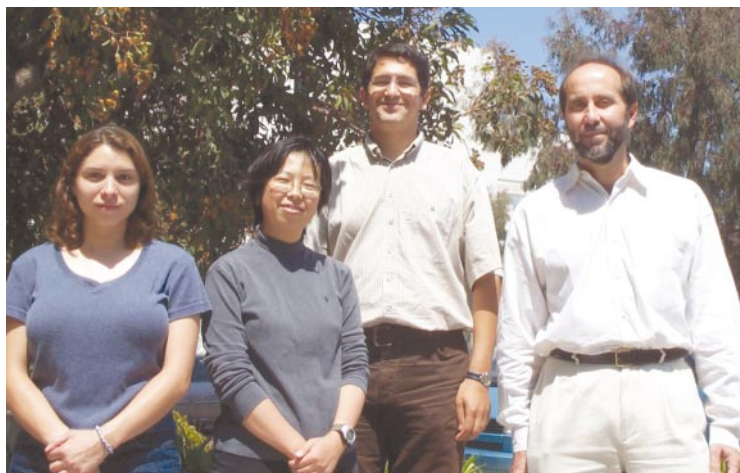
This approach is catching on — Harvard University's Bauer Center for Genomics Research, which opened last year, takes a similar tack. It is an angle that attracted Aviv Regev, a computational biologist who recently took up a fellowship at the centre. She strongly agrees that studying both benchtop biology and computational science contributed to her own success. Theoretical students need to learn first-hand how experimental biologists solve problems, she adds.

"Your ongoing science is going to be composed of moving between two very different systems," says Regev, who says she constantly switches between thinking intuitively about a biological problem and fashioning it and a possible solution into a computational problem with known parameters.

It is not only life scientists who are seeking new partnerships to help them to tackle complexity; physical scientists also increasingly seem to be looking for ways to branch out.

This trend is reflected in recent meetings. In

José Onuchic (right) and his team work at the interface between theory and experiment.



September 2002, the American Physical Society held a meeting entitled 'Opportunities in Biology for Physicists' that attracted 200 young physicists. In Britain, the Engineering and Physical Sciences Research Council created the Life Sciences Interface in 1999. The programme funds large group grants of £10 million (US\$16 million) for medical imaging, bio-nanotechnology and tissue engineering, and facilitates cross-talk workshops and informal week-long meetings that throw biologists and physicists together to see if they can come up with new ways to solve biological puzzles.

"One of the things most often given as a barrier to working at the interface is time," explains John Hand, the programme's director. "So we created the Discipline Hopping scheme." This offers individual researchers up to £50,000 to venture into another field by taking a sabbatical, travelling or hiring a cross-discipline research assistant for the lab.

Scientists who can move fluently between disciplines may end up with double the academic opportunities: biology departments are searching for bioinformatics specialists and computational biologists, whereas physics departments are looking to expand into biological physics.

COMPUTING POWER

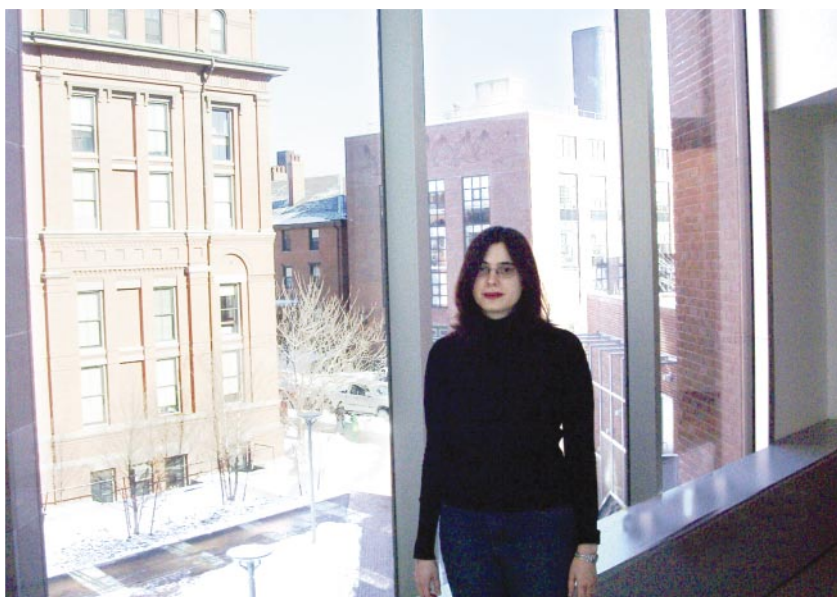
If any one group within biophysics is set to cash in on the genome, it is computational biologists. Demand for them is skyrocketing as the number of sequenced genomes rises. According to most researchers in genomics, a severe shortage of computational biologists remains the biggest bottleneck in the task of turning sequenced genomes into meaningful interpretations.

"Unfortunately, there are not enough people who have both an understanding of biology experiments and of the algorithms involved in bioinformatics," says Lakshmanan Iyer, a computational biologist at Harvard who has provided technical support for bioinformatics research at both Harvard Medical School and the Bauer Center for Genomics Research. Iyer trains scientists to use web-based sequence-analysis tools and also advises on individual project strategies.

"If you have a specialization in statistics or maths with exposure to biology, then you are really set," he says. The International Society for Computational Biology's website job list reflects the wealth of opportunities, with almost 40 academic, research institute or genetic-based biotechnology positions posted in the past three months.

Although many computational biologists are being snatched up to work on the genome level, others are concentrating on the next level of integration, the pathway. Systems biology has got off to a flying start as the research community has come to realize that understanding the 'interactome' — all of the possible protein-protein interactions for a cell — will be the next big hurdle in biology.

To prepare for the challenge, systems biologists are beginning to form coalitions and consortia — much like in the early days of the Human Genome Project. The Alliance for Cellular Signaling has a US\$25-million National Institute of General Medical Sciences (NIGMS) grant as well as corporate sponsorship. It



Aviv Regev believes that a flexible approach is essential for computational biology

plans to track all signalling events in mouse immune and heart-muscle cells.

Other efforts at integration include BioPAX, a public database for keeping track of known pathway information, and a new computer modelling language, Systems Biology Markup Language, or SBML, that will allow researchers directly to compare computational models of biological systems.

Regev looks forward to the day when her models can be matched up easily with others. She is applying new forms of calculus, originally designed for mobile-phone networks, to cellular signal-transduction cascades. The phone networks, she says, are analogous to biological systems, where molecular interactions happen concurrently rather than linearly and determine "who speaks to whom". She is betting that quantitative modelling of well-known pathways such as the MAP kinase cascade will yield better predictions of interactions.

STRUCTURAL BOOM

But before jumping from completed sequence to complete systems, there is a big bridge to be built, and with significant labour from biophysicists: solving the 'protein structure genome'.

This is the field of structural genomics — the marriage of genomics, computation and experiment that aims to speed up the determination of structures of most of the proteins whose structure and function are still unknown. Ambitious structural-genomics initiatives are under way in the United States, Germany and Japan, in particular, and promise to increase the workload at synchrotron X-ray and nuclear-magnetic-resonance (NMR) facilities. The European Union's Sixth Framework Programme has designated E1.1 billion (US\$1.2 billion) for bioinformatics and structural-genomics work from 2002 to 2006.

"We need to solve 20,000 structures to characterize about 90% of the human genome," says Andrzej Joachimiak, director of the Structural Biology Center at Argonne National Laboratory's Advanced Photon



Nuclear-magnetic-resonance spectrophotometers, such as this 800-MHz machine, are key to elucidating protein structures.

Source (APS) in Illinois. He and others say that it will take the concerted effort of many scientists trained in structural biology, including those with skills in protein purification and crystallization, to complete this task. Joachimiak is also principal investigator for the Midwest Center for Structural Genomics, one of nine structural-genomics consortia supported by the Protein Structure Initiative, a NIGMS programme.

With a projected half-a-billion dollars in funding, the initiative seeks to add 10,000 new unique structures to the Protein Data Bank in ten years. Ideally, that would result in one representative protein structure for each protein family,

which would serve as templates for solving other members of the families.

"It could create a catalogue for a short cut to all of the protein structures on the planet," says Charles Edmonds, a programme director at the NIGMS. As such, the nine centres are not limited to human protein structures; their subjects range from plants and nematodes to the bacteria that cause tuberculosis.

But to accomplish this feat, each centre must significantly ramp up its production in the next two years to churn out 200–300 structures each year. Last year Joachimiak's centre solved 45. To boost output, the NIGMS, together with the National Cancer Institute, has agreed to build three new synchrotron X-ray beamlines at the APS at a total cost of about \$20 million.

The beamlines provide high-brilliance X rays that can resolve large and complex molecular structures in a fraction of the time needed with conventional X-ray sources. For example,

Joachimiak and his colleagues solved the structure of a small bacterial protein in under 7 hours. More typically, he says, you can go from crystal to structure in about 30 hours for proteins that might have taken three years with earlier methods.

The advent of structural-genomics initiatives has raised the demand for beamtime at existing facilities, prompting the construction of new synchrotrons and additional beamlines at existing ones. Gerhard Materlik, chief executive of Diamond, the new UK synchrotron scheduled to open in early 2007 in Chilton, says that about half of the first-phase beamlines will be dedicated to life-science research. He anticipates hiring 200–250 technical staff by 2010.

"We're really looking for qualified experts," he says. "There is not an abundance of people who can run these machines." Competition will be fierce — other projects scheduled to come online include the Synchrotron Light Laboratory in Barcelona, the Australian Synchrotron in Melbourne, and the Canadian Light Source in Saskatoon. These facilities require physicists to build, develop and maintain the beamlines, crystallographers to advise users and help to collect data, and administrators to make decisions that will guarantee efficiency. Even well-funded facilities have trouble keeping up with demand and finding enough staff to run experiments around the clock.

"I compare these facilities to driving a very sporty car, like a Ferrari, 24 hours a day, seven days a week," says Joachimiak. Running state-of-the-art technology can mean fast-paced excitement and constant problem-solving challenges, but it can also easily lead to burnout when helping hundreds of users each year.

Managers agree that these positions, although comparable in salary to an assistant-professor position, can be frustrating for someone who wants to develop their own research programme. Most of the job is devoted to helping users run experiments and trouble-shooting. To keep highly trained

A joint effort

Wah-Keat Lee spends 80% of his working week helping researchers to use an X-ray synchrotron beamline at the Argonne National Laboratory's Advanced Photon Source in Illinois. But he still manages to fit in time for some research of his own — and his latest project has just landed him a front cover of *Science*.

As part of his work, Lee helps to test new X-ray optics. One day, he picked up a dead ant from the floor and put it in the path of the X-ray beamline. He was amazed to see the insect's internal organs rendered in incredibly fine detail.

He investigated the technique in collaboration with researchers from the Field Museum of Natural History in Chicago, and the team last month published a paper documenting a

previously unknown breathing mechanism in ants, crickets and beetles (M. W. Westneat *et al. Science* 299, 558–560; 2003).

Lee has also used the method to study teeth in sea urchins, fuel-injection gas nozzles and crack formation in aluminium construction materials. "It dramatically improves the contrast in the image and brings out features that would otherwise not be seen," Lee says.

If he can find enough interested users, Lee would like to optimize the beam for imaging live animals so that organs in zebrafish larvae, or even anaesthetized mice, could be resolved to a thousandth of a millimetre.

Lee's success is an example of a growing trend in which technical staff are encouraged to participate in research collaborations.

Anthony Watts, a biophysicist at the University of Oxford, UK, points to the successful research networks set up among the major European neutron facilities. These networks bring technical staff into collaborations designed to improve the instrumentation or measurement methodologies. Watts attributes the success of the policy to the fact that the workers "get integrated into the research projects they are meant to be supporting".

Paul Ellis, who oversees the macromolecular-structure group at the Pacific Northwest National Laboratory in Richland, Washington, agrees that the involvement of technical staff is crucial. "We stress here that they are not just a pair of hands," he says. "They have to be involved in the collaboration intimately."

K.P.

technical staff stimulated, some facilities are encouraging ongoing research collaborations for their workers (see 'A joint effort', opposite). And, Joachimiak points out, unlike an academic position, facility funding is extremely stable and doesn't require grant writing.

The cheaper and faster structure determination becomes, the more attractive it is to pharmaceutical companies. Ten major companies share a beamline at the APS for high-throughput screening of hundreds of drug candidates bound to their target protein. And the drug-discovery company Structural GenomiX, based in La Jolla, California, built its own beamline at the APS for a gamble on structure-based drug discovery.

A German initiative — the Protein Structure Factory — hopes to crank out human protein structures based on the German human genome project. This high-throughput assembly line starts with complementary DNA sequences and ends with crystal-structure determination done at the BESSY II synchrotron in Berlin. These large-scale endeavours will need biology-literate computer scientists who can optimize the beamline system for automation.

MAGNETIC MOMENTS

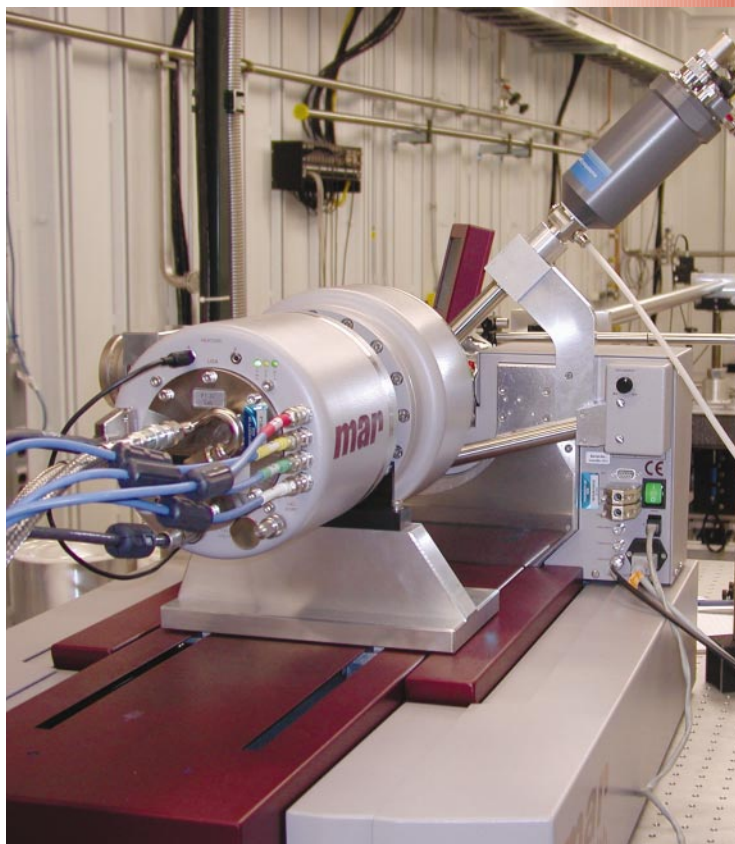
Another technological development that will make an important contribution to the structural-genomics initiative is the use of high-field NMR spectrophotometers. These machines use a powerful 900-MHz magnet to reduce analysis time and boost resolution. They can be used to tease apart large protein complexes or to examine protein–protein and protein–DNA interactions. "It brings experiments into reality that would have otherwise been shelved," says Paul Ellis, director of the macromolecular structure and dynamics group at the Pacific Northwest National Laboratory in Richland, Washington.

But such facilities are still relatively rare — there are just two in the United States, at Pacific Northwest and the Scripps Research Institute in La Jolla. This should soon change, as the National Institutes of Health has recently awarded grants for four machines to the Massachusetts Institute of Technology in Cambridge, the New York Structural Biology Center in New York City, the University of Georgia in Athens, and the University of Wisconsin, Madison.

Mastering technology outside a given researcher's discipline is an important step towards broader career options. Anthony Watts, who directs the National Biological Solid State NMR Facility at the University of Oxford, UK, encourages flexibility among his staff. "I try to expose all of the people to as many technologies as possible," he says. "If they want to be independent in academia or industry, they need to take on as many tasks as possible."

Watts' lab has an 800-MHz NMR spectrophotometer, which he uses to study membrane proteins that might serve as drug receptors. He says that physicists in his lab learn how to culture bacteria, whereas the biologists find out how to design NMR experiments.

Demand within the marketplace for those who master the range of technical skills is strong. "Technical staff is a seller's market," says Edmonds. "There are just not enough people who have a spectrum of



skills that spans a detailed understanding of computer-controlled instrumentation and the details of experimental structural biology."

This is especially good news for anyone with dual training or who wants to move from physics or biology into biophysics. A survey by the US Biophysical Society of biophysics students and postdocs in 1999 showed that 84% believed they would land a position in their chosen field — incredible optimism at a time when the National Research Council reported that, for the life sciences in general, there were too many people being trained.

Their ability to sift through stacks of data quantitatively and arrive at biologically relevant answers looks set to provide biophysicists with a secure future in the biology lab. "There is an enormous amount of data that has to be turned into a picture — there are a lot of opportunities out there," says Iyer. ■

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

Web links

Center for Theoretical Biological Physics

♦ ctbp.ucsd.edu

Bauer Center for Genomics Research

♦ cgr.harvard.edu

Advanced Photon Source

♦ www.aps.anl.gov

Pacific Northwest National Laboratory

♦ www.pnl.gov

Protein Structure Factory

♦ www.proteinstrukturfabrik.de

Diamond

♦ www.diamond.ac.uk

Online: drug-discovery firm Structural GenomiX has its own X-ray beamline at the Advanced Photon Source in Argonne.